DELINEATING THE MOLECULAR INTERACTIONS OF GRO CHEMOKINES

Ph.D THESIS

by

KHUSHBOO GULATI

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667, INDIA
AUGUST, 2017

DELINEATING THE MOLECULAR INTERACTIONS OF GRO CHEMOKINES

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

KHUSHBOO GULATI

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667, INDIA
AUGUST, 2017

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**DELINEATING THE MOLECULAR INTERACTIONS OF GRO CHEMOKINES**", in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2013 to August, 2017, under the supervision of Dr. Krishna Mohan Poluri, Assistant Professor, Department of Biotechnolgy, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

(**KHUSHBOO GULATI**)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(KRISHNA MOHAN POLURI)
Supervisor

The Ph. D. Viva-Voce examination of **Ms. Khushboo Gulati**, Research Scholar has been held on ……..

Chairman, SRC **Signature of External Examiner**

This is to certify that student has made all the corrections in the thesis.

Signature of Supervisor **Head of the Department**

Dated: _____

Table of Contents

Chapter 2

Mechanistic Insights into Molecular Evolution of Species Specific Differential Glycosaminoglycan Binding Surfaces in GRO Chemokines

Chapter 3

Deciphering the Oligomerization and GAG Binding Characteristics of CXCL1 and CXCL2 Chemokines

"ISHWAR YAT KAROTI SHOBHNAM KAROTI"

ACKNOWLEDGEMENTS

First of all, I shall pay obeisance in the name of the Almighty God and express my thanks and profound gratefulness to him for showering his blessings and strength that made me competent to amble on the avenue leading to the success of my life.

I take immense pleasure in expressing my deep sense of gratitude to my supervisor Assistant Professor Dr. Krishna Mohan Poluri for his unwavering support, collegiality inspiration, timely guidance, constructive criticism, profound concern, and mentorship throughout my Ph.D. It was a great privilege and honor to work and study under his precious guidance. His dynamism, vision, sincerity, and motivation have deeply inspired me. I express my wholehearted indebtedness to him.

I am thankful to Dr Dinesh Kumar (CBMR Lucknow), for his valuable teachings and providing the NMR facility at CBMR Lucknow. I am also very thankful to Dr. Sebastian Köhing, university of Berlin, Germany for providing synthetic GAGs.

Besides, I would like to thank the Department of Biotechnology, and Institute Instrumentation Centre (IIC), IIT Roorkee for instrument facilities to carry out the research work. I am also grateful to my SRC committee members: Dr Pravindra Kumar, Prof. R.P. Singh, and Dr. Rama Krishna Peddinti; DRC chairman: Prof. Vikas Pruthi; Head of the Department: Prof. Partha Roy; for their continuous advice. I am also very thankful to Prof. Ritu Barthwal, Dr. Sulakshana P. Mukherjee, Dr. Debabrata Sircar, Dr. Gopinath and Prof. Partha Roy for their timely and kind suggestions. I also express my sincere gratitude towards the other faculty members of Department of Biotechnology and all my teachers to guide me throughout my education.

Nevertheless, I am thankful to all of my lab mates and colleagues Meenakshi Sharma, Krishnakant Gangele, Nidhi Joshi, Nipanshu Agarwal, Mukesh Kumar Meher, Jaya Lakshmi, Apporva Goel, Neha Arora, Zia Tariq, Nancy Jaiswal (CBMR, Lucknow),Mohd. Yousuf, Neelanchal Vaid, for providing a stimulating and creative environment that greatly assisted my research. In addition, I would like to specially thank master's student Minal Jamsandekar who worked with me and gave me an opportunity to learn new things from her. I have great pleasure in acknowledging my gratitude to seniors like Sweta Tripathi, Pradeep TP, Padmapriya, for their kind help and support.

I greatly acknowledge Prof. Daniel Huster (Leipzig University, Germany) for allowing me to work in his lab. I am also very thankful to Peter Schmidt, Lars, Cindy, Matthias, Holger, Anja, Alexender, Ulrike, Juliane, Anika, Marcel, Mario, Vaidehi, Rewati, Venus, and Supriyo, for their kind support during my stay in Leipzig, Germany.

I am very much thankful to Prof Krishna Rajarathnam, from Department of Biochemistry and Molecular Biology, The University of Texas Medical Branch (UTMB), Galveston, Texas, USA, for kindly providing the permission for reference NMR spectrum of CXCL1.

I am thankful to Prof. Neel Sarovar Bhavesh from International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India for providing the access to their NMR facility.

My time at IITR was made enjoyable due to friends like Ankita, Lipi, Preeti, Pooja, Nishu, Anjali. A special note of thanks to my off campus friends like Gaurav Rose (Gauri), Supriya Sharma (Supu), Jaspreet Aujla (Jassi), Vandana Sandhu (Raavan), Bharti (Olu), Jaspreet Kaur Rai (Banvari), and Shruti Dhingra, for their encouragement and support.

I also thank the office staff, Department of Biotechnology, campus administration and workers team of my hostel, canteen, Sarojini bhawan, their hard work made my stay comfortable in campus.

I gratefully acknowledge Ministry of Human Research and Development (MHRD) the funding source that made my Ph.D. work possible. Furthermore I would also like to acknowledge Science and Engineering Research Board (SERB) for providing me the fellowship to attend XXVIIth ICMRBS-2016, at Kyoto, Japan. I also thank ICMRBS-2016 committee for the young investigator award. I am really very much thankful to Leipzig University for visiting scholar fellowship (2017) to support my lab visit at Leipzig, Germany.

Last but not a least, I would like to express my deepest gratitude to my beloved grand parents (Lt Sh. Ram Pyara Lal Gulati and Lt Smt. Vidya Gulati), and my family members for supporting me spiritually throughout my Ph. D and my life in general.

Abstract

Chemokines are small 8-10 kD signaling entities that are involved in numerous physiological and pathological processes including leukocyte trafficking, angiogenesis, organogenesis, tissue development, and tumorigenesis etc, through their association with cell surface glycosaminoglycans (GAGs) and the G-protein coupled receptors (GPCRs). Humans express around 50 chemokines which are segregated into different classes namely CXC, CC, CX3C, and C chemokines based on the positioning of N-terminus cysteine residues. Chemokine monomers share the canonical tertiary structure comprising of long unstructured N-terminus, followed by 3_{10}-helix, three antiparallel β-sheets, and a C-terminal α-helix.

Chemokines oligomerize into dimers, tetramers, and other higher order oligomers. They form two types of dimers, namely CC and CXC. CXC dimer is more globular and is formed by the two antiparallel α-helices lying on the top of six stranded antiparallel β-sheets, whereas CC dimers involve the formation of a two stranded antiparallel β-sheet between the N-terminal regions of the two monomeric units. Recent studies have also reported the formation of heterodimers between several CC/CXC chemokines thus adding complexity to the underlying process of chemokine regulated leukocyte trafficking processes.

Neutrophil activating chemokines (NACs) belong to the subfamily of CXC chemokines which are specifically involved in recruitment of neutrophils. NAC family comprises of seven proteins namely CXCL1, CXCL2, CXCL3, CXCL5, CXCL6, CXCL7 and CXCL8, with a signature ELR (Glutamic acid- Leucine- Arginine) motif at their N-terminus. Among NACs, CXCL1 (GRO α), CXCL2 (GRO β), and CXCL3 (GRO γ) are called as Growth Related Oncogene chemokines (GRO). GRO family arose as a result of two rounds of gene duplication during the course of evolution. GRO chemokines are closely related to each other, and are involved in growth and progression of melanoma tumors. Despite their super-close relativity in sequence and structure, biological studies have reported the differential expression patterns of all three GRO genes in tissue and signal specific manner, and their involvement in different functions. They have also adapted different mechanisms and signaling pathways to accomplish their specific functions. The differential behaviors of these highly related members of GRO chemokines can be implicated to their differential evolution patterns, oligomerization (homo/hetero), and variable interactions with cellular partners (GAGs and receptors). Although plethora of literature is available about the differential function of GRO proteins, however, the mechanistic details of their evolution and

structure-function relationships at molecular level are scarce. Hence, the research work in the current thesis has been designed to decipher the evolutionary perspectives, oligomerization potencies, differential GAG binding interactions, and structure-stability relationships of GRO chemokines at atomic level by using multitude of biophysical and computational techniques. Specific details of thesis chapters (1 to 4) are as follows.

Chapter 1 provides an overview of chemokine research area including the role of chemokines in innate and adaptive immunity, origin of chemokine family, classification of chemokines, their structural and other biological properties including oligomerization, interactions with chemokine receptors and GAGs. A detailed literature survey of different types of NACs, their biological importance, structures solved by X-ray crystallography and NMR spectroscopy, studies focused on interaction of NACs with cellular binding partners including receptor and GAGs have also been discussed with special emphasis on known literature of GRO chemokines. This chapter also provides the snapshot of various bioinformatics and biophysical tools used in the thesis to accomplish the proposed objectives.

Chapter 2 presents the evolution-structure relationships of GRO chemokines. The evolutionary mechanisms including selection pressures involved in diversification of GRO genes have been characterized by performing comprehensive evolutionary analysis among different mammalian species. Phylogenetic analysis showed species specific evolution pattern in GRO chemokines, while the selection analysis revealed that these genes have undergone concerted evolution and possess positive selection sites, although majority of them are under purifying selection. Interestingly, positively selected sites are more concentrated on the C-terminal / GAG binding and dimerization segment. Substitution rate analysis also corroborated the species specific evolution, and confirmed C-terminal domain of GRO genes as the highest substituted segment. Further, the structural analysis established that these nucleotide alterations in the GAG binding domain are the source of surface charge modulation, thus generating the differential GAG binding surfaces and multiple binding sites as per evolutionary pressure, even though the helical surface is primordial for GAG binding. Indeed, such variable electrostatic surfaces are crucial to regulate chemokine gradient formation during host's defense against pathogens thus explaining the significance of chemokine promiscuity.

Chapter 3 characterizes the homo / hetero oligomerization potencies and GAG binding features of CXCL1 and CXCL2. Genes for murine CXCL1 and CXCL2 were successfully cloned, expressed, and proteins were purified using the combination of various chromatography

techniques. Homo-oligomerization propensity of CXCL1 and CXCL2 assessed by size exclusion chromatography (SEC) and glutaraldehyde cross linking assay demonstrated the strong dimeric nature of CXCL2 as compared to CXCL1. NMR ^{1}H-^{15}N HSQC experiment of CXCL1 and CXCL2 at same concentration (~ 150 μM) evidenced two set of peaks corresponding to both monomer and dimer for CXCL1, whereas a single set of dimeric resonances in CXCL2 spectrum establishing their intrinsic differential homo dimerization capabilities. To further characterize the resonances from the monomeric and/or dimeric species and dimer interface residues, NH cross peaks in the ^{1}H-^{15}N HSQC spectra of both CXCL1 and CXCL2 were assigned. The formation of CXCL1/CXCL2 heterodimer was characterized experimentally by mixing their equimolar concentrations and monitoring the ^{1}H-^{15}N resonances of CXCL1. The addition of unlabeled CXCL2 resulted in new set of dimer-interface peaks along with the existing two sets of resonances (CXCL1 monomer/homo-dimer), accompanied by a significant attenuation in the peak intensities. Quantification of intensities of the dimer interface residues Q25 and L30 of CXCL1 before and after the addition of CXCL2 indicated the presence of ~ 40 % heterodimer under chosen experimental conditions, thus directly demonstrating the potential formation of CXCL1-CXCL2 heterodimers.

Upon assessing the oligomerization characteristics of CXCL1/2 homo/hetero dimers, different oligomeric species of these chemokines were subjected to interactions with GAGs/ GAG mimetics (HP6-heparin hexasaccharide, HA6-hyaluronan hexasaccharide, SHA6-synthetic sulfated hyaluronan hexasaccharide, and NC6-neocarradodecaose hexasulfate) in order to monitor the phenomenon of GAG induced oligomerization. NMR titration experiments indicated that HP6 and SHA6 induced dimerization in CXCL1, CXCL2, CXCL1/2 heterodimer in contrast to HA6 and NC6. These results indicated that sulfation of GAGs indeed is essential to oligomerize/dimerize chemokines. Moreover, the results evidenced that the extent and positioning of the sulfation also play a crucial role in regulating the chemokine GAG interactions and GAG induced chemokine oligomerization.

Chapter 4 depicts the biophysical characterization of CXCL3 and its comparative analysis of the structure-stability-dynamics relationship with CXCL2. Murine CXCL3 gene was successfully cloned, expressed, and purified partly as soluble form (supernatant) from cytoplasm and rest as insoluble form from inclusion bodies to maximize the protein yield. Resonance assignments of CXCL3 NH cross peaks were obtained using conventional 3D-NMR experiments. Oligomerization state of CXCL3 appraised using size exclusion chromatography, 2D-DOSY, and

^{15}N-NOESY indicated that the dimeric nature of CXCL3 is similar to CXCL2. Heparin binding assay suggested the GAG binding affinity of CXCL3 and CXCL2 are weaker compared to CXCL1. Temperature dependent Circular Dichroism (CD) studies evidenced that both the CXCL2 and CXCL3 structures are highly stable to thermal perturbations. Structural alignment and contact map analysis of these two chemokines implied their structural equivalence with differential electrostatic surface potentials on both β-sheet and α-helix surfaces. Further, tertiary structural characteristics of CXCL3 and CXCL2 investigated using ANS fluorescence showed distinct fluorescence profiles. Binding mode of ANS to CXCL3 and CXCL2 explored using fluorescence life time analysis (FLS) and NMR suggested the presence of surface exposed specific hydrophobic binding pocket on helical surface of CXCL3 in contrast to CXCL2.

To further gain insights into their residue-level differential stability-dynamics properties of CXCL2 and CXCL3, NMR based native state hydrogen exchange (NHX) studies, temperature dependence NMR and ^{15}N backbone relaxation studies were performed. Stabilization free energies evidenced that CXCL2 is comparatively more stable than CXCL3. Temperature dependence, amide proton chemical shifts, and ^{15}N relaxation studies unraveled the differential dynamic features of these two paralogs at both faster (ps-ns) and slower (μs-ms) time scales. These differential structural, stability and dynamic features demonstrate that although CXCL2 and CXCL3 share similar tertiary structural and oligomerization features, they do possess specific electrostatic surfaces, varied dynamics and stability characteristics that can be implicated to their differential / specific functional behaviors.

In summary, the present thesis unraveled the evolutionary, structural, and functional relationship of GRO chemokines and characterized their structural, dynamic, oligomerization, and GAG binding features at residue level, thus providing mechanistic insights into their molecular interactions. This comprehensive study will aid chemokine/GAG researchers in rational designing of novel chemokine and/or GAG mimetics to regulate neutrophil trafficking.

List of Publications

1. **Gulati K**, and Poluri KM "Deciphering the in vitro homo and hetero oligomerization characteristics of CXCL1/CXCL2 chemokines" RSC Adv., 2016, 6:28213–28218.
2. **Gulati K**, Jamsandekar M, and Poluri KM "Mechanistic Insights into Molecular Evolution of Species Specific Differential Glycosaminoglycan Binding Surfaces in GRO Chemokines" R. Soc. open sci., 2017, 4(9): 171059.
3. **Gulati K**, Gangele K, Agarwal N, Jamsandekar M, Kumar D, and Poluri KM "Molecular Cloning and Biophysical Characterization of CXCL3 Chemokine" Int J Biol Macromol., doi: 10.1016/j.ijbiomac.2017.09.032, (in press).
4. **Gulati K**, and Poluri KM "Mechanistic and therapeutic overview of glycosaminoglycans: the unsung heroes of biomolecular signaling" Glycoconj. J, 2016, 33:1–17.
5. **Gulati K**, and Poluri KM "Chemoattractants, Scaffolds and Endogenous Stem Cells: Adorable Partners of In Situ Tissue Regeneration", Austin J Biotechnol Bioeng, 2015, 2(4):1052.
6. **Gulati K**, Meher MK, Poluri KM. "Glycosaminoglycan-based resorbable polymer composites in tissue refurbishment" Regen Med, 2017, 12(4):431-457
7. Poluri KM, and **Gulati K**. "Protein Engineering techniques Gateways to synthetic Protein Universe" Springer Brief, (Book), ISBN 978-981-10-2732-1
8. **Gulati K**, and Poluri KM "An Overview of Computational and Experimental Methods for Designing Novel Proteins" Recent Pat Biotechnol., 2016, 10(3):235-263.
9. Arora N, **Gulati K**, Alok Patel A, Parul A. Pruthi, Krishna Mohan Poluri KM, Pruthi V. "A hybrid approach integrating arsenic detoxification with biodiesel production using oleaginous microalgae" Algal Research, 2017, 24 (Part A):29–39.
10. Rao ABP, **Gulati K**, Joshi N, Deba DK, Rambabuc D, Kaminskyd W, Poluri KM, Kollipara MR. "Synthesis and biological studies of ruthenium, rhodium and iridium metal complexes with pyrazole-based ligands displaying unpredicted bonding modes" Inorganica Chimica Acta, 2017, 462(1):223–235.

List of Conference Presentations

Oral presentation:

1. **Gulati K**, Gangele G, and Poluri KM "Investigations on Evolution-Structure Epitome of Glycosaminoglycans Binding to Dimeric CXCL1" during the "Recent Trends in Chemical Science, Engineering and Technology (RTCET) 2014 at NIT, Hamirpur, in May 2014.

Poster Presentations:

1. **Gulati K**, and Poluri KM "Deciphering the Evolution-Structure relationship of Growth Related Oncogene Chemokines" during the Annual Conference of Indian Biophysical Society entitled "Molecular Architecture, Dynamics and Assem" (MADALS) 2014 at Saha institute of Nuclear Physics, Kolkata, in Feb 2014.
2. **Gulati K**, and Poluri KM "Mechanistic insights into the oligomerization and differential GAG binding potencies of Gro α Chemokine" during the 22nd Conference of NMR Society of India (NMRS) 2016 at IIT Kharagpur, in Feb 2016.
3. **Gulati K**, and Poluri KM "Mechanistic insights into the oligomerization and differential GAG binding potencies of Gro α Chemokine" during the Biotech day, at Department of Biotechnology, IIT Roorkee in Feb 2016.
4. **Gulati K**, and Poluri KM "Dissecting the oligomerization features of murine Chemokines KC and MIP2" during the XXVIIth International Conference on Magnetic Resonance in Biological Systems (ICMRBS) 2016 at Kyoto international conference center, Kyoto, Japan, in Aug 2016.

Chapter 1: Introduction to Chemokine Biology and Biophysical Techniques

1.1 Immune system

Immune system is the host's own defense mechanism, which protects their body against various environmental and pathogenic insults. It counter-attacks and combats with harmful pathogens via two types of immune functionalities. They include (a) innate immunity (also known as humoral immunity) and (b) adaptive immunity (also known as cell mediated immunity). Innate immunity provides the first line of defense against the infections, which implies that its combat action starts with immediate effect, i.e., just after or during the onset of progression of the pathogen invasion/infection. The foremost element participating in the initiation of innate immune response involves the recognition of components of invading pathogen, known as pathogen associated molecular patterns (PAMPs), which includes microbial metabolic products such as LPS, peptidoglycans, lipoproteins un-methylated DNA containing CpG motif. These PAMPS are recognized by the PAMP receptors such as toll like receptors that releases anti-inflammatory molecules (cytokines, chemokines etc.) for self-defense [1]. In contrast, the adaptive immune system comes into action after five or six days of infection and is highly specific in identifying pathogens. Moreover, the adaptive immune system is responsible for the memory associated immune response. It is also a heightened immune response against the pathogen for its neutralization and clearance when the exposure to the same antigen is repeated. Major players in the adaptive immune response include; lymphocytes, (majorly B-cells and T-cells), and the molecules including growth factors cytokines, and antibodies, that they secrete.

Both adaptive immune system and innate immune system are not independent of each other. Phagocytes involved in the non-specific immune response ultimately triggers the specific immune response by activating the cells of the adaptive immune system, which in turn secrete molecules like cytokines to further enhance the action of the phagocytes. Thus, both the adaptive and innate immune system works in a coordinate manner to combat the infectious agents. In general, the innate immune responses are able to clear most of the infections from the body prior to

their activation of the adaptive immune system [2]. All the immune actions like antigen presentation, cell-cell recognition, and activation of other cells are mediated by various immune system proteins. There is a large diversity among the proteins which allows them to recognize a diverse variety of cell types, and to respond to infinite antigenic challenges. These proteins either belong to the category of membrane proteins (toll like receptor, B7 family of co-stimulatory proteins, B-cell receptor, T-cell receptor, MHC Class I and MHC class II molecules, cell surface immunoglobulins, cytokine and growth factor receptors etc.) or the extracellular secretory proteins like cytokines, antibodies, and chemokines.

Numerous proteins participating in the immune system are highly related, and are the translation products of the multigene families. Multigene families are the group of genes that are highly similar to each other and thought to have arisen from the common ancestor. These groups of highly related genes are a direct consequence of duplication events, followed by numerous other evolutionary events including mutations, polymorphisms, exon shuffling, splicing, and pseudogenization etc. Several members of multigene families are highly related to each other, some are highly distant, whereas some previously existing members have now lost during the evolutionary selection processes. Such selection criteria followed by nature, have resulted in a diverse variety of genes that are related to each other in several aspects. They also have acquired novel functions, as they are under tremendous pressure to counter attack the microbes that are developing novel and rapid mechanisms to exploit the host immune machinery for their own survival. On the basis of such relationships among evolved genes/proteins, several evolutionary models have been proposed for the origin of the multigene families, namely concerted gene evolution, divergent evolution, and birth and death evolution. These models have been followed by the numerous immune system related genes including Immunoglobulins [3], major histocompatibility complex (MHC) [4], cytokines, chemokines, and disease-resistance genes [5]. Among various immune related gene families, cytokines and chemokines belongs to the category of most rapidly evolving genes within mammals [6].

1.2 Cytokines

Cytokines are small 5-20 kD signaling proteins that are sub-divided into interleukins, interferons, tumor necrosis factors, and chemokines. They function in autocrine, paracrine or endocrine manner thereby act as immune modulators. Interleukins including IL-1, 2, 3, 4, 5, 6, 7,

8, 9, 10, 11, and 12 primarily targets leukocytes and are involved in their growth, development, and differentiation. Interferon family with three members including IFN-α, IFN-β, and IFN-γ exhibits antiviral activity which inhibits the viral replication in infected cells and also protects the uninfected cells from the infection, by stimulating natural killer cells and cytotoxic lymphocytes. Tumor necrosis factor includes two homologous proteins namely; TNF α and TNF β, which are involved in creating antitumor immunity by the exertion of their cytotoxic effects on cancerous cells and promotion of antitumor immune responses. TNF α plays an important role by regulating the early production of chemokines and endothelial adhesion molecules including endothelial leukocyte adhesion molecule (ELAM I), which is highly effective in neutrophil binding to endothelial cells and for neutrophil recruitment [7]. Chemokines are the chemotactic cytokines which are involved in migration of various immune cells along the chemokine gradient. Other cytokines includes colony stimulating factors namely G-CSF, GM-CSF, and M-CSF that provides allergic immunity. Anti-inflammatory cytokine like TGF-β regulates the cell growth and exhibit both stimulatory and inhibitory effect on various cell types [8-10]. Cytokines interact with their cognate receptors and mediate intracellular signaling processes.

1.3 Chemokine gene family origin

Chemokines belong to the multi-gene family encompassing numerous chemokines and chemokine receptor members that are involved in diversified functions. Like other multigene families, chemokine family is also originated from various duplication events by modification/alteration of chemokine genes to modify or acquire novel functions. Evolutionary studies have reported that the chemokine genes have evolved from the single ancestor, 650Mya, through number of duplication events and are still processing [11-13]. Studies involving the genomic arrangement of chemokines by the syntenic analysis across the genomes of mammalian species suggested that the chemokine family expanded mainly through the contribution of tandem gene duplication events and numerous evolutionary processes such as birth and death, insertion/deletion, alteration of nucleotides etc. [13,14].

1.4 Classification of chemokines

Around 50 chemokines and 22 chemokine receptor proteins are expressed in humans. These small 8-10 kD cytokine proteins have been segregated on structural grounds into four

different classes namely, CXC, CC, CX3C and C chemokines, based on the arrangement and number of Cys residues present at the N-terminal [6]. In CXC chemokines, two Cys residues are separated by the presence of one amino acid 'X', whereas in CC chemokines, two Cys residues are adjacent to each other. CX3C class comprises of three amino acids 'XXX' between the two Cys residues. C/XC chemokine members contains only one cys residue at N-terminal **(Fig.1.1)**.

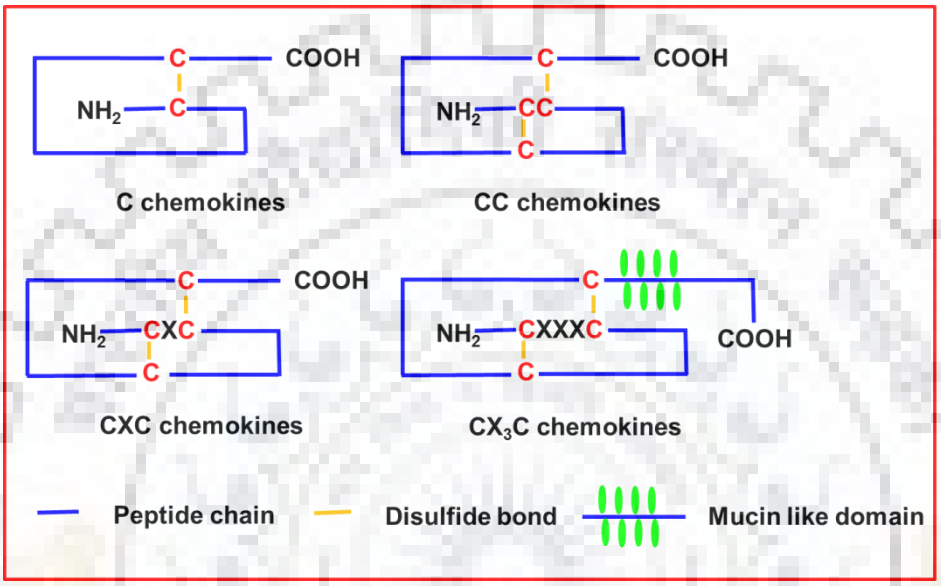


Figure 1.1: *Arrangement of disulfide bonds (yellow) in different classes of chemokines. C represents Cys residue and X can be any other amino acid.*

Chemokines have also been divided into different classes on the basis of their basic functionalities. They include: inflammatory chemokines, homeostatic chemokines and dual chemokines **(Fig. 1.2)**. Inflammatory chemokines gets activated upon an inflammatory response, and are mainly involved in leukocyte trafficking. Homeostatic chemokines are the constitutive chemokines that always remain expressed in the lymphoid or other organs, and are involved in migration and homing of numerous different kinds of cells including lymphocytes. Dual chemokines which exhibits overlapping functions of both inflammatory and homeostatic chemokines. Some inflammatory chemokines remains constitutively expressed and performs homeostatic functions, whereas some homeostatic chemokines becomes hyper active under extreme conditions [15].

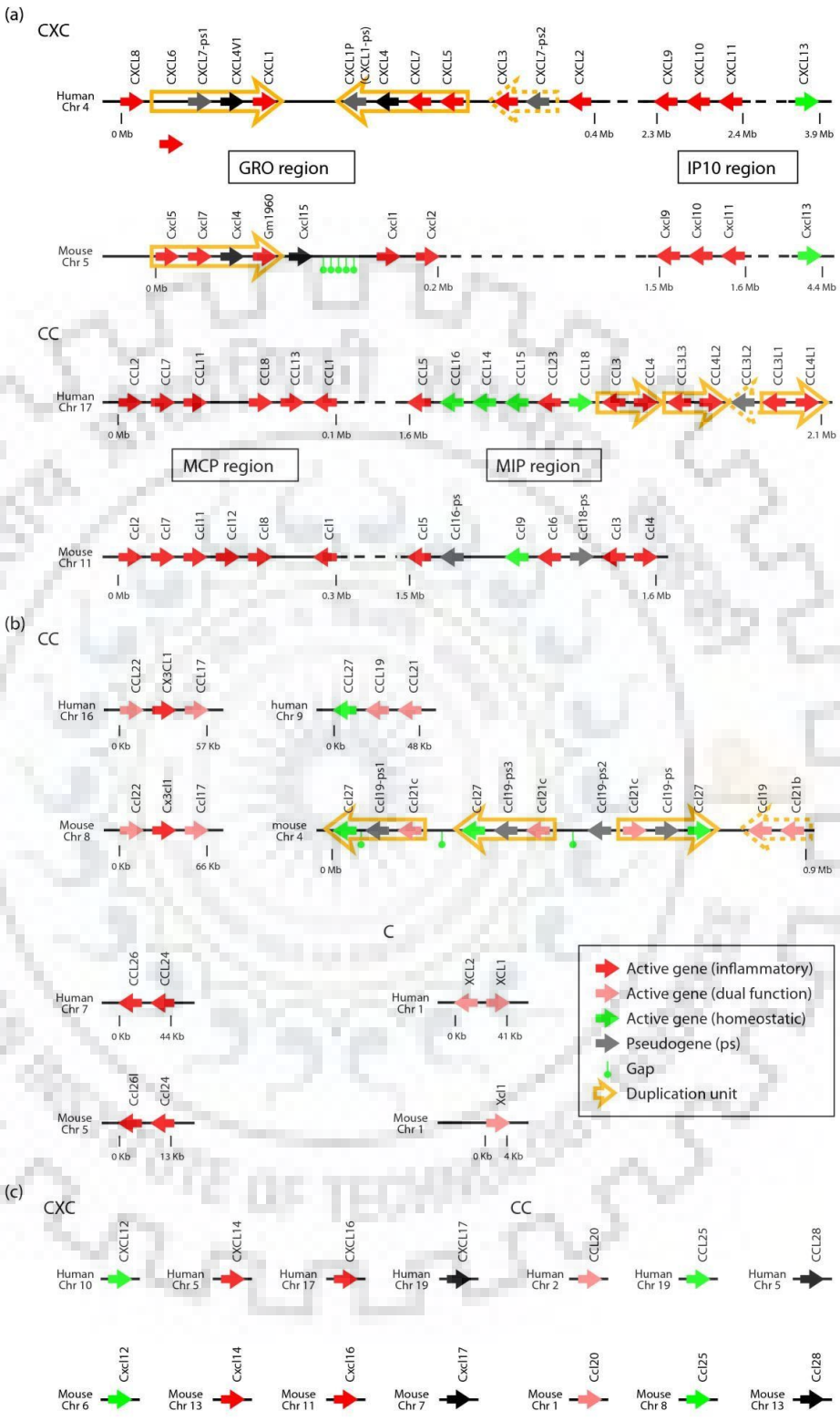


Figure 1.2: *Chromosomal arrangement of chemokines in humans and mouse:* **(a)** *major cluster,* **(b)** *minor cluster,* **(c)** *non-cluster chemokines. Solid arrows indicate chemokine genes and their transcriptional orientation, red, green, and black represent the inflammatory, homeostatic and*

dual function chemokines respectively and gray indicates pseudogenes. Open yellow arrows in the major cluster indicate duplication units. Gap indicates the region not covered by genome sequencing consortiums, while dashed line indicates a similar region of more than 1Mb. Figure adapted and reproduced from Zlotnik et al 2006 [6]. Copyright under CCA Licence 4.0.

Further, chemokines have also been sectioned into different groups based on their genomic arrangement on chromosomes **(Fig 1.2)**. They are: (a) major cluster chemokines, whose genes are present in large clusters on chromosome; and (b) mini-cluster or non-cluster chemokines, whose genes are present at discrete locations on chromosome. There are two major clusters of CXC chemokines and CC chemokines, along with several other mini-clusters. This clustered arrangement of chemokines is the resultant of duplication of genes happened during the course of evolution. These cluster chemokines perform the common primary function along with their own specific functions. The two major clusters in CXC chemokines are: (a) GRO (growth related oncogene)/NACs (neutrophil activating chemokine) and (b) IP10 (Interferon-gamma inducible protein). In CC chemokines, they are (a) MCP (monocyte chemotactic protein) and (b) MIP (macrophage inflammatory protein) clusters. Due to the clustering of chemokines, they also show promiscuity with different receptors, which implies that multiple chemokines can bind to a single receptor and vice versa [6,16,17].

1.5 Chemokine receptors

Chemokine receptors (CXCR, CCR, CXC3CR and CR) have been classified on the basis of chemokines to which they bind [18]. They belong to G-protein coupled receptor (GPCR) family of receptors. Various sets of chemokine receptors are expressed on different types of leukocytes to facilitate the interaction for specific chemokine ligands. Chemokine receptors interact with their respective ligands, and undergo conformational changes to activate intracellular effectors, initiate signaling pathways thus resulting in various cellular responses. Chemokines interact with the receptor using a two site mechanism in which one site involves the binding of N-terminal region of receptor into the core domain of chemokine which is formed by N-loop and β_3 strand of chemokine and the other site involves the binding of N-terminus of chemokine to the ligand binding pocket of the receptor. The two steps of interaction occur sequentially, contributing to the binding of chemokine to the receptor and subsequently, the activation of the receptor [19,20]. Some of the chemokine receptors are not GPCRs but are atypical receptors (CXCR7, CCBP2, CCRL1, CCRL2, and DARC). These atypical chemokine receptors are also known as non-

chemotactic, scavengers, or recycling receptors. They majorly function as chemokine transporters, scavengers or decoy receptors [21,22]. Presence of unequal number of chemokine and chemokine receptors itself indicates that there is no one to one relationship applicable for all the chemokines and chemokine receptors. Many chemokines can bind to a single receptor and similarly a single chemokine can bind to more than one receptor suggesting that there is promiscuity in their binding relationship **(Fig. 1.3)** [23]. However, the specificity in their mechanism of action has been reported as different chemokines binding to the same receptor regulates different signaling pathways [24]. This selectivity among chemokine and chemokine receptor pairs indicate the fine tuning of the chemokine/chemokine receptor mediated cell signaling responses against different stimuli [23,25].

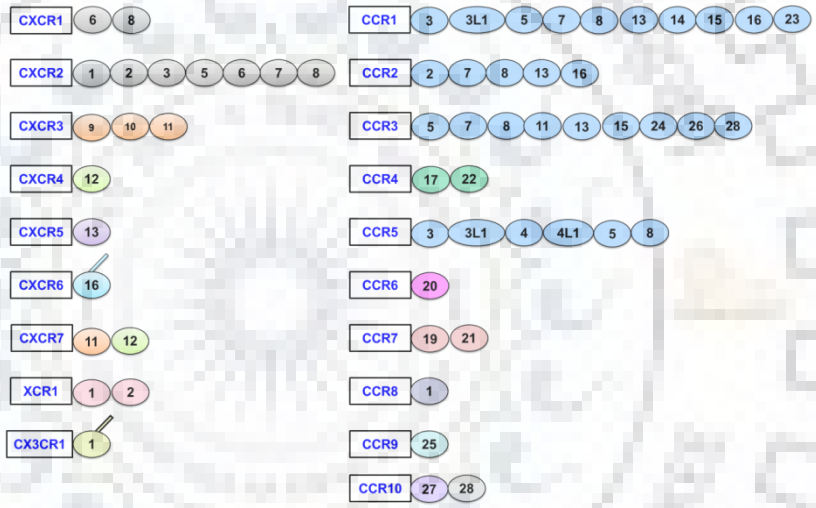


Figure 1.3: *Promiscuous relationship of chemokines and chemokine receptors. Boxes and circles indicate the chemokine receptor and chemokine ligands to which they bind respectively. Same color of the circles indicates that they lie in the same chromosome. Extra line in CXCL16 and CX3CL shows that they exist as transmembrane proteins.*

1.6 Chemokines mediate leukocyte trafficking to resolve inflammation

Release of cytokines by the interaction of PAMPs with PRRs amplifies the release of chemokines to initiate the process of inflammation. These chemokines causes the rapid recruitment of leukocytes at the site of infection to marshal the infectious agent. The recruitment process involves; first the release of chemokines from tissue macrophages, and the formation of their gradient on the luminal surface of endothelial cells as a result of their interaction with endothelial cell surface GAGs. This is then followed by the initial attachment and rolling of leukocytes on the

endothelium. The event of attachment and rolling is mediated by the interaction of P, E selectins present on active endothelial cells and L-selectins, being constitutively expressed on leukocytes. Chemokines present on endothelium interacts with chemokine receptors present on leukocytes results in integrin activation and firm adhesion of these cells. Leukocytes keep on sensing the chemokine gradient and migrate towards the site of infection by undergoing the events of shape change, including the formation of their arms and legs by the polymerization and breakdown of actin leading to the formation and retraction of lamellopodia, and extravasation through endothelium [26]. Leukocytes finally, kill the pathogens either by phagocytosis, degranulation or by using their extracellular traps **(Fig. 1.4)**. All these events clearly mark the importance of chemokines which regulates and provides the guidelines for the controlled process of leukocyte trafficking, which is an important component of innate immune response [26,27].

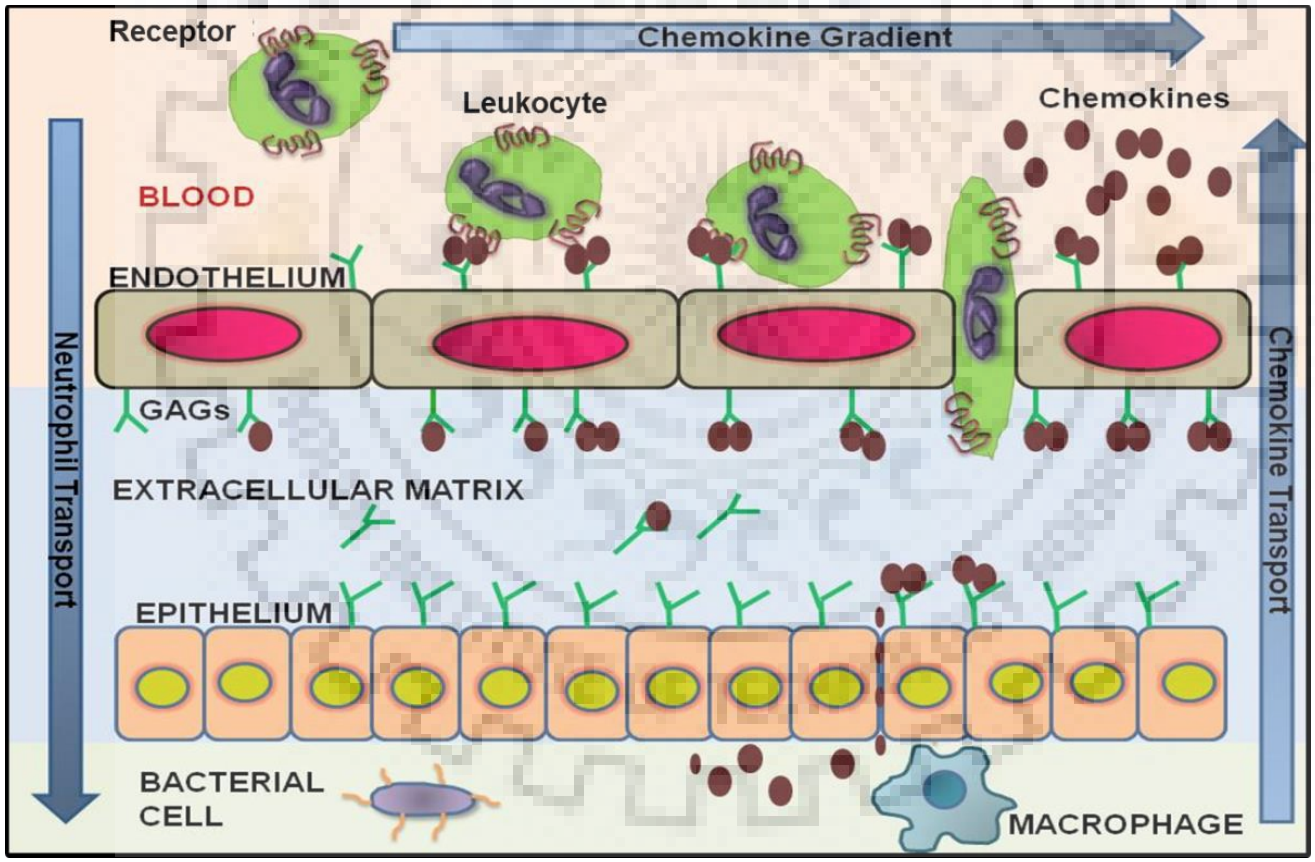


Figure 1.4: *Schematic showing the various steps involved in chemokine mediated leukocyte recruitment.*

1.7 Other biological/immune regulatory functions of chemokines

Chemokines in addition to their role in leukocyte migration for immune surveillance, they also plays a significant role in migration of cells during other cellular processes including angiogenesis, organogenesis, cell proliferation, tissue development, wound healing, tumorigenesis, and neural regeneration etc. **(Fig. 1.5)** [28,29]. They are also involved in controlled migration, homing, and maturation of various cells of adaptive immune responses including T- and B-lymphocytes. Thus, chemokines mediate the coordinated migration and interaction of immune cells of innate and adaptive immune responses thereby, meeting the requirement of both the innate and adaptive immunity [28,30-32]. Chemokine receptor axis is also associated with numerous diseases due to the deregulation of the cellular processes carried out by them [33,34]. Chemokines are key mediators in pathogenesis of chronic inflammatory diseases like atherosclerosis, rheumatoid arthritis, cardiovascular diseases [35-37]. Specific chemokines and receptors play key roles in progression of cancers, as they cause metastasis of cancer cells, and also aids in shaping the tumor microenvironment by leukocyte recruitment and activation of pro-inflammatory mediators [38]. Chemokine receptors (CXCR4 and CCR5) also serve as potential entry sites for the HIV to evade the immune system [39,40]. The involvement of chemokine and chemokine receptors in pathogenesis of such diversified diseases highlight their potentials to serve as attractive drug targets [41]. Insights into different pathological and physiological function of chemokines and their receptors help in designing various chemokine and chemokine receptor based therapeutics against various diseases.

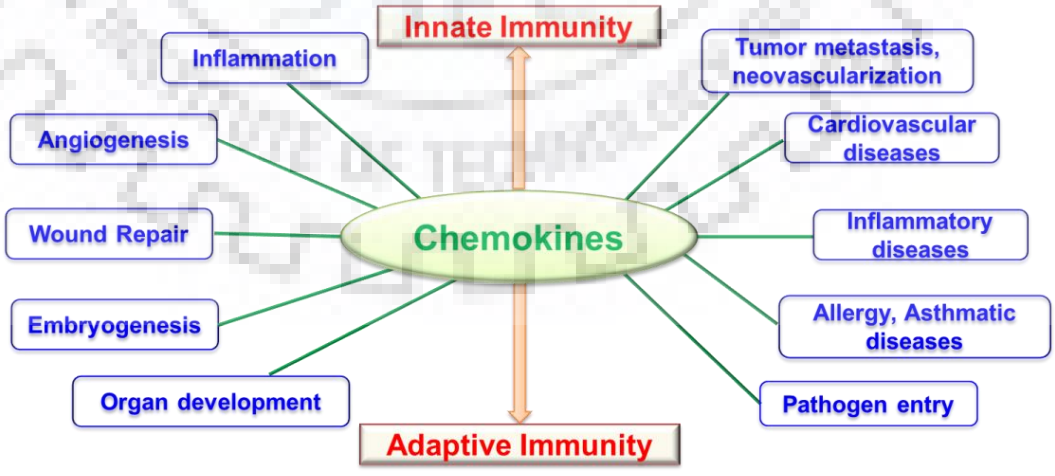


Figure 1.5: *Various physiological and pathological roles of chemokines.*

1.8 Structural characteristics of chemokines

Despite of different classes and differential roles of chemokines, all of them share the canonical tertiary structure [42]. The signature fold of monomeric chemokines includes long unstructured N-terminus, followed by 3_{10}-helix, three antiparallel stranded β-sheet and a C-terminal α-helix **(Fig. 1.6)**. Structure of the chemokine is stabilized by the formation of disulfide bond(s), which connects the N-terminus to the core structure of chemokine. Chemokines possess various loops to connect different secondary structural elements including: 'N-loop' is an unstructured that connects the N-terminal to the first β-strand, '30S-loop' connecting the first and second β-strands, '40S-loop' connects second and third β-strands, and '50S-loop' connects the third β-strand to C-terminal α-helix **(Fig. 1.6)**. Studies reported that chemokines involves the 3_{10} and C-terminal α-helix to interact with GAGs and their large unstructured N-terminal domain interacts with their cognate G-protein coupled receptors **(Fig. 1.6)**.

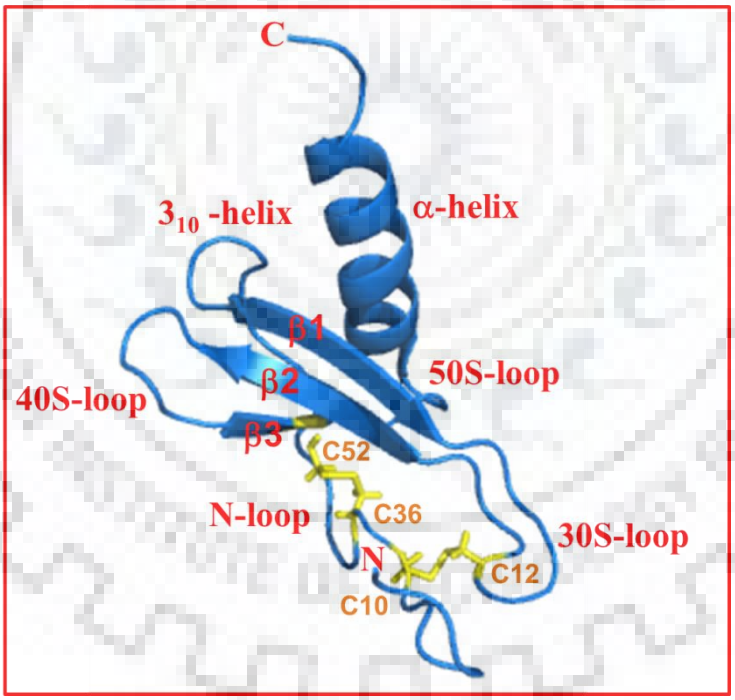


Figure 1.6: *Basic topology of chemokines represented by the monomeric structure of human CXCL1 (PDB ID: 1MGS). Essential structural elements comprises of 2 disulfide bonds (in yellow), different loops (N-loop, 30S, 40S and 50S loops) connecting three β-sheets, 3_{10} helix, and C-terminal α helix.*

1.9 Homo and hetero oligomerization of chemokines

Oligomerization is an important phenomenon, which is found in ~ 70% of eukaryotic proteins [43,44]. Oligomerization provides multitudes of benefits to the proteins including: (1) induces the stability to the proteins and protects them from proteolytic actions, (2) amplifies their efficiencies in signaling events or in their enzymatic actions, (3) increases the avidities of proteins for their ligands by assembling into polymeric constructs or arrays, (4) adding complexity and more diversity towards the regulation of signaling processes [26,31,45].

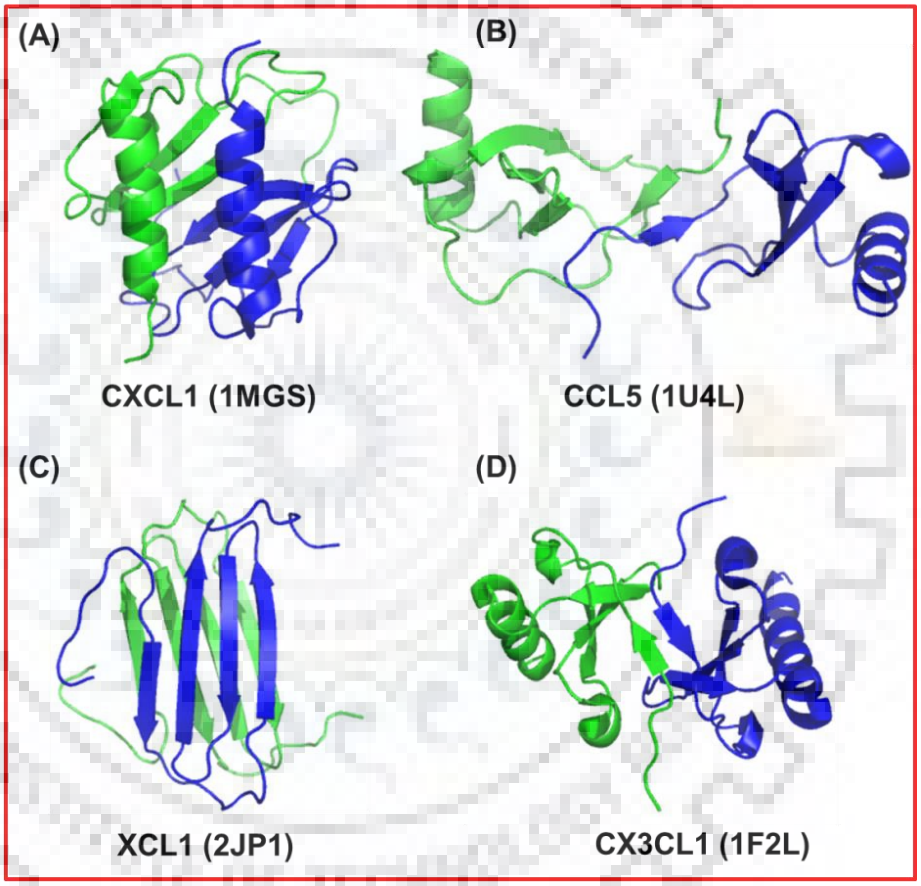


Figure 1.7: *Different type of dimeric structure formed by different classes of chemokines:* **(A)** *CXC type dimer,* **(B)** *CC type dimer,* **(c)** *XC type dimer,* **(d)** *CX3CL type dimer. PDB IDs for each structure are enclosed in parenthesis.*

Chemokines exhibit both homo and hetero oligomerization potency, indicating that they can form dimers, tetramers, even other higher order oligomers. CXC and CC chemokines even though share the same monomeric fold, they form different dimeric structures. CXC chemokines involves their first β-strand from each monomeric unit, resulting in the formation of six stranded

antiparallel β-sheet topped by the two C-terminal α-helices **(Fig. 1.7 A)**. This type of dimer arrangement is known as CXC-type dimer, which leaves the N-terminal, N-loop and β_3-strand free to be accessible for interaction with receptor **(Fig. 1.7 A)**. CC chemokines forms CC type dimers, which involve the formation of a two stranded antiparallel β-sheet between the N-terminal regions of two monomeric units to form an elongated dumbbell shape dimer **(Fig. 1.7 B)**. The monomeric subunits in the dimer are stabilized by various interactions including hydrogen bonding, electrostatic, and hydrophobic interactions. In CC-type dimers, the involvement of N-terminal region in the dimer formation makes them inaccessible for its interaction with receptors, thus essentially inactivating the dimer conformation. CX_3C and XC chemokines also forms dimers that are different from CXC and XC type dimers. XC chemokines/lymphotactins exist in an equilibrium of two conformational states. One is the monomeric form which exhibits the canonical tertiary fold of chemokines and other is dimeric form with all beta sheet structure **(Fig. 1.7 C) [44]**. CX3C is a unique chemokine class that contains only one member fractalkine (CX3CL1). CX3CL forms CC type dimer but the core of the monomeric units are closer to each other as compared to normal CC type dimer **(Fig. 1.7 D)**. This is due to the unique positioning of disulfide bonds in CX3CL [46].

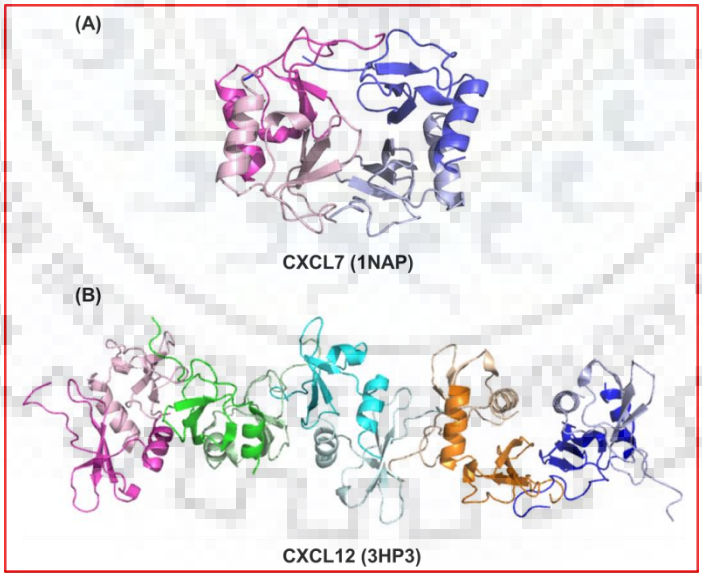


Figure 1.8: *Higher order oligomers in chemokines represented by CXCL7 (tetramer) and CXCL12 (decamer). PDB IDs for each structure are enclosed in parenthesis.*

Higher order oligomers of chemokines have also been captured either alone or in conjunction with GAGs [47]. Tetramerization has been observed in crystal structure of CXCL7 in

which two CXC type dimers of CXCL7 are arranged back to back **(Fig.1.8 A)** [48]. Decameric structure of CXCL12, formed by the five dimers of CXCL12, in which monomeric units are arranged in the dimer using CC-type dimer interface, and monomeric units of adjacent dimers interact uniquely by involving the C-terminal helix from the one monomer and the first β-strand from the other monomer [49] **(Fig. 1.8 B)**. The oligomerization phenomenon in the chemokine family generates diversified conformations, thus contributing to their variable functional characteristics [50].

Chemokines also undergo hetero oligomerization. Two types of hetero oligomers have been reported. (1) Hetero oligomers formed between the chemokines that belong to the same family, for example, CCL3-CCL4, CXCL1-CXCL7. (2) hetero oligomers formed between the chemokines that belongs to different family, for instance, CXCL4-CCL5, CXCL4-CCL2 [51-54]. Heterodimers including CCL3-CCL4 and CXCL4-CCL5 have been reported to perform synergistic actions [53,54]. Both homo and hetero oligomers are targeted as therapeutic agents [55].

1.10 Chemokine interaction with GAGs

Glycosaminoglycans (GAGs) are long, linear polysaccharides that are expressed on cell surfaces and found in extracellular matrices. GAGs are classically divided into four families considering their chemical composition. They are; heparin/heparan sulfate (HS), chondroitin/dermatan sulfate (CS/DS), keratan sulfate (KS), and hyaluronan (HA). GAGs are composed of the repeating disaccharide unit containing hexuronic acid that can be either D-glucuronic acid (D-GlcA) or L-Iduronic acid (L-IdoA) (with exception of KS containing galactose instead of uronic acid) and hexosamine includes N-acetyl-D-glucosamine (D-GlcNAc) or N-acetyl-D-galactosamine (D-GalNAc). HA, Heparin and HS containing glucosamine are called as glucosaminoglycans, whereas the rest having galactosamine are known as galactosaminoglycans. Except hyaluronan (HA), other GAGs are highly negative charged molecules **(Fig. 1.9)**.

GAGs being the negatively charged entities interact with proteins through electrostatic interactions, and thereby act as repository for the highly diffusible proteins thus present them to their receptors as and when required. GAGs interact with numerous of proteins including growth factors, cytokines, and chemokines to mediate different cellular functions. The interaction of proteins with GAGs plays an important role in regulating the cellular processes. Chemokines with

highly basic nature, interact with GAGs, and are immobilized by GAGs on cell surfaces for their presentation to the receptors [56,57].

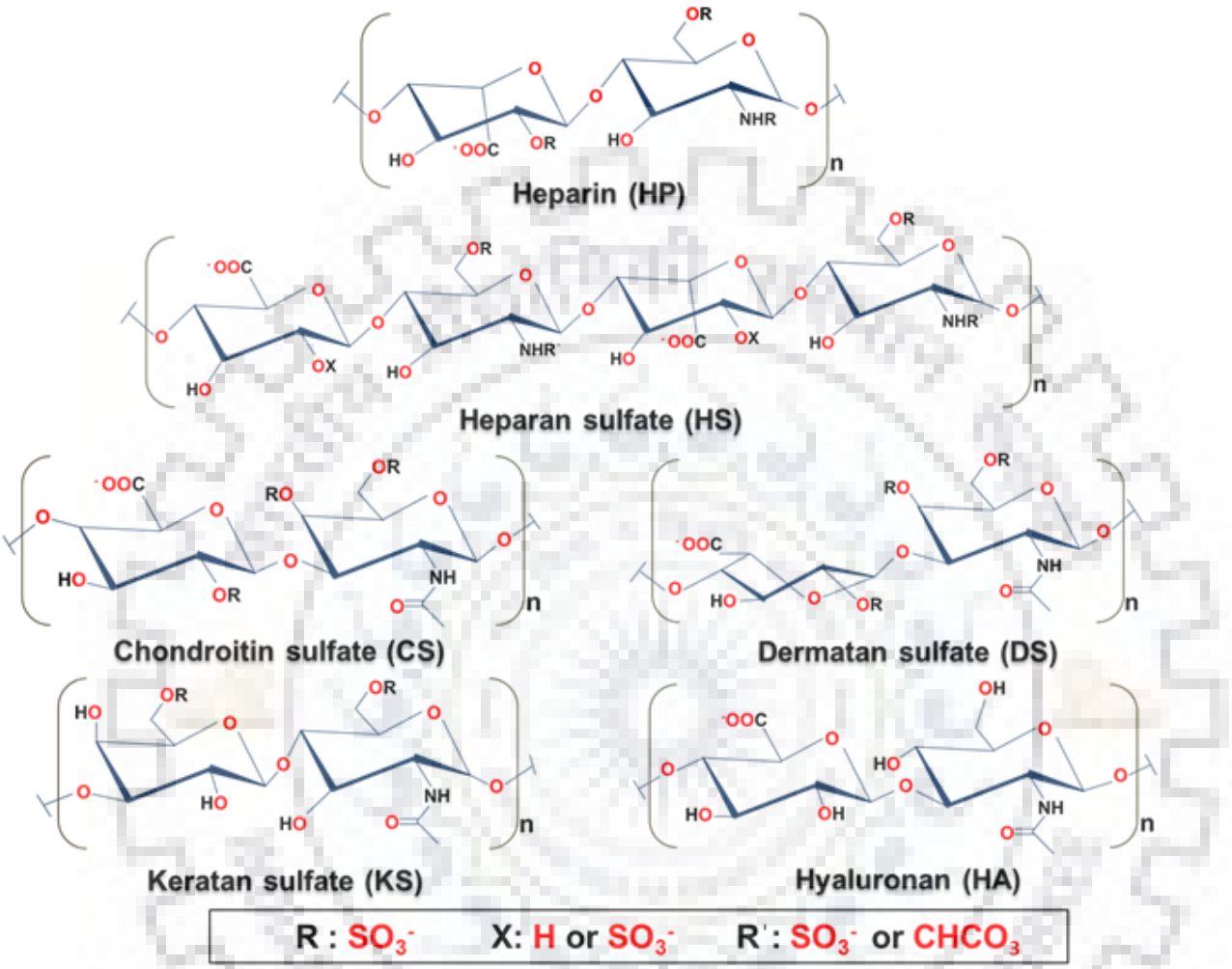


Figure 1.9: *Repeating disaccharide/tetrasaccharide units of different glycosaminoglycans. Heparin disaccharide is shown using its major disaccharide of L-iduronic acid and D-glucosamine. Tetrasaccharide repeating unit of heparin sulfate with two distinct disaccharide units: one unit composed of D-glucuronic acid and D-glucosamine, and the other with L-iduronic acid and D-glucosamine.*

Numerous other roles are also being played by chemokine-GAG interactions in addition to the localization of chemokines for the formation of chemokine gradients under sheer force of blood flow to provide directional clues for cell migration *[58]*. The additional roles include: (a) GAGs protect the chemokines from proteolytic degradation, and thereby regulate their presentation on endothelial cell surfaces for receptor interactions [59-61]; (b) GAGs protect the chemokines from their inappropriate involvement and interaction with receptors [62]. (c) Chemokine-GAG

interactions also promote chemokine secretions from T-cells, tumors cells, facilitate the transcytosis of chemokines across the cells, increase their stability, and also induces signaling pathways [63-67].

Giving due importance to biological functions carried out by chemokine-GAG interactions, studies are being carried out to unravel the specificities and molecular details of these interactions. Several mutagenesis, structural, and biophysical studies proved that chemokine-GAG interactions are highly specific and selective towards GAGs [68-74]. Specificity of chemokines (CCL5, CCL3, CCL2, and CXCL8/IL-8) towards different GAGs including heparin, heparan sulfate (HS), dermatan sulfate (DS), and chondroitin sulfate (CS) were explored [75,76].

GAG binding epitopes of chemokines have also been probed in several of the chemokines. Kuschert et al were the first to entail the GAG binding residues for IL-8/CXCL8 and established that residues (K60, K64, R68) present at C-terminal helix contributes majorly for the heparin binding [77]. Residues Lys-58 and His-66 in the C-terminal alpha-helix of MCP-1 are reported to play an important role in GAG binding [78]. These findings indicated the C-terminal helix serves as the GAG binding domain which is located away from the receptor binding domain. These reports lead to the hypothesis that chemokine can bind both GAG and receptor simultaneously. However, latter studies indicated the presence of GAG binding residues at N-terminus were overlapping with the receptor binding residues. It is not yet clear that chemokines have the ability to bind GAGs and receptors simultaneously or sequentially [79-81]. Characterization of other chemokines including CCL3(MIP1α), CCl4(MIP1β), and CCL5(RANTES) revealed the presence of BBXB motif in 40S loop which is essential for the GAG binding [82-85]. The BBXB motif is also present in CXC chemokines CXCL12, and CXCL10 in other structural segments [86,87]. Altogether, these studies indicate the specificity of chemokines for GAGs with a distinct distribution of GAG binding epitopes.

1.11 Chemokine based therapeutics

Chemokines with diversified physiological roles including hematopoiesis, angiogenesis, and organogenesis are being exploited for the development of new organs, tissue regeneration, and other biomedical applications [88,89]. Novel chemokine based fusion toxin protein (F49A-FTP) has been rationally designed as an antiviral therapeutic by the fusion of F49A variant of CX3CL1 with the catalytic domain of Pseudomonas exotoxin A (PE) [90]. Specific chemokines such as

CXCL12 either alone or along with other chemoattaractants including growth factors are being delivered at the injury site to increase their local concentration to promote the recruitment of endogenous stem cells during in situ tissue regeneration **(Fig. 1.10)** [91,92]. Other chemokines like IL-8 (CXCL8), MCP1 (CCL2), MIP-1 (CCL3), RANTES (CCL5), TARC (CCL17), SLC (CCL21) and MDC (CCL22) have also been employed effectively in mesenchymal stem cell (MSC) recruitment and tissue regeneration [93].

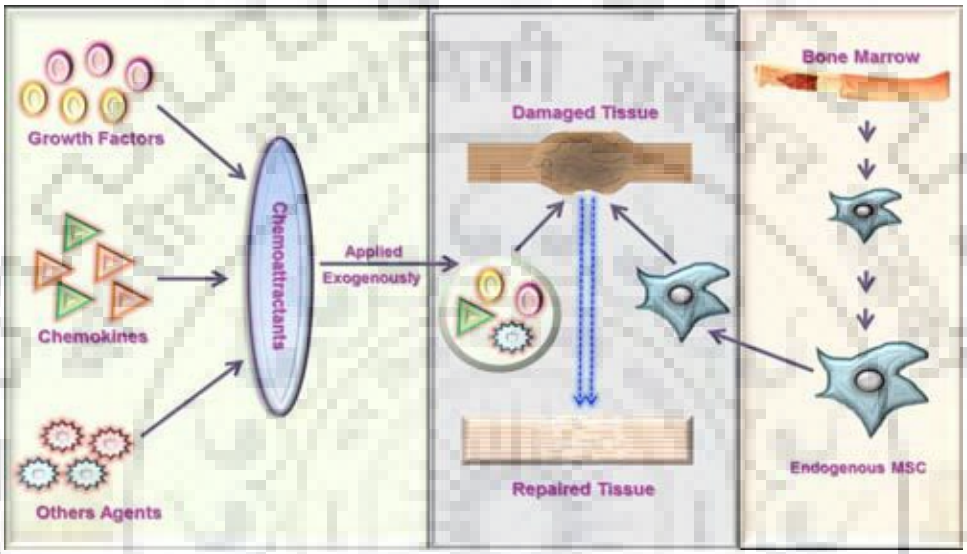


Figure 1.10: *Schematic showing the migration and recruitment of endogenous MSCs by the exogenous chemoattractants including chemokines, growth factors and other agents enveloped in an engineered scaffold at the site of injury for tissue regeneration/repair.*

Chemokines inherent properties including their short half-life, amenability for cleavage by proteases, rapid diffusion by bolus infections and inflammatory side effects makes them a tough choice for therapeutics. Hence, it is essential to use an appropriate delivery device for delivering chemokine based drugs. Variety of devices have been developed for the effective release of chemokines including self-assembling peptides, covalent binding to PEGylated fibrin patch, heparinized collagen scaffold, mineralized collagen type 1 scaffold, poly (lactic-co-glycolic acid) (PLGA), poly (ε-caprolactone), poly (lactideethylene oxide fumarate) hydrogel and chitosan/poly(gamma-glutamic acid) complexes etc. [91]. The glycosaminoglycan (GAG) based fabrications in the form of heparin/hyaluronic acid based hydrogels, heparinized collagen scaffolds, heparin-coating, and poly (L-lysine)–hyaluronan multilayer films are being successfully employed for delivery of chemokines for various biomedical applications [93]. For example, polymeric disks containing GAGs have been devised for the delivery of chemokine analogue 5P12

RANTES for its prolonged and sustained delivery for the HIV prevention [94]. Pharmaceutical composition for such polymer bonded sulfated GAG based chemokine delivery device has been recently patented [95].

GAG based chemokine delivery devices are not only designed for delivery of chemokines, but also for combined release of GAGs and chemokines to exert synergistic effects on cell homing and migration. For instance, hyaluronic acid based hydrogels with degradable crosslinks, has been designed for the delivery of CXCL12 and hyaluronic acid in the injured myocardium to increase the bone marrow derived cell (BMC) homing for myocardium remodeling, which served better as compared to the delivery of CXCL12 alone [96]. Such scaffolds add another layer of advantage for chemokine delivery as they mimic natural GAG-chemokine interactions and also efficiently protect them from proteolytic degradation [97].

1.12 Neutrophil activating chemokines

Neutrophil activating chemokines (NACs) belong to the subfamily of CXC chemokines that are involved in recruitment of neutrophils to the site of infection. NACs play an important role in providing first line of defense. NAC family comprise of seven members including CXCL1, CXCL2, CXCL3, CXCL5, CXCL6, CXCL7 and CXCL8. They comprise of conserved ELR (Glutamic acid- Leucine- Arginine) motif just prior to the first Cys at the N-terminus **(Fig. 1.11)**. CXCL1, CXCL2, and CXCL3 are known as Growth related oncogene (GRO) chemokines; hence are also represented as GROα, GROβ, and GROγ respectively. CXCL5, CXCL6, CXCL7, CXCL8 are known as ENA78 (epithelial-derived neutrophil-activating peptide 78), GCP2 (granulocyte chemoattractant protein-2), NAP2 (Neutrophil activating peptide2) and IL-8 (Interleukin-8) respectively. GRO genes are 90% identical in amino acid sequence, whereas CXCL5, CXCL6, CXCL7, and CXCL8 are 40-50% identical to each other and also with GRO proteins. The presence of ELR motif confers the specificity of these chemokines to the CXC chemokine receptors (CXCR1 and CXCR2) present on neutrophils [11,97]. NACs are secreted by numerous cell types in response to various stimuli including pro-inflammatory cytokines like IL-1 and tumor necrosis factor (TNF).

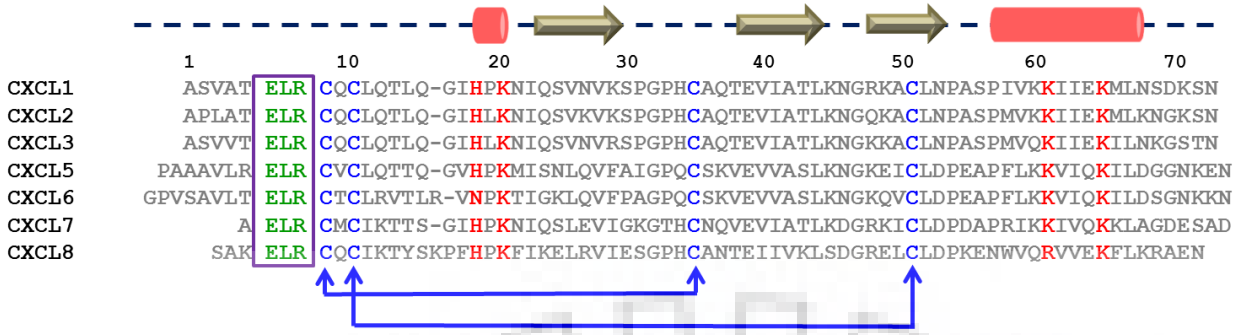


Figure 1.11: *Sequence alignment of human neutrophil activating chemokines. Conserved ELR motif essential for receptor binding is marked in green and conserved cysteine residues for disulfide formation is highlighted in blue, and residues important for GAG binding are marked in red. Numbers are marked with respect to CXCL1 sequence.*

1.13 Assessment of NAC functions

NACs are involved in angiogenesis and chemotaxis of endothelial cells [8]. CXCL1, CXCL2, and CXCL3 serve as activating factor for basophils, eosinophils, smooth muscle cells, and lymphocytes which implicates their role in acute inflammation [98]. CXCL5 (ENA78) is expressed by the highly specialized cells including white adipose tissues macrophages in muscles, cardiomyocytes, and alveolar epithelial type II cells [99,100]. CXCL5, in adipose tissues, acts as an adipokine by activating the JAK2/STAT5 pathway, thereby blocks insulin signaling and promotes obesity [100]. In lung, CXCL5 interacts with CXCR2 to mediate neutrophil trafficking in response to bacterial infection by activating both G-protein and arrestin signaling pathways [99,101]. CXCL5 is also associated with numerous acute and chronic diseases including obesity, rheumatoid arthritis, inflammatory bowel disease, sun burn, and different cancers [102-105]. CXCL6 (GCP2) is expressed by numerous of the epithelial cells including airways, eyes, gastrointestinal tract, mammary glands, tonsils, macrophages and mesenchymal cells during inflammation [106]. In the absence of inflammation, it is expressed in various types of cancers including non-small cell lung cancer [107], breast cancer [108], colorectal cancer, and endometrial cancer [109]. CXCL6 has also been shown to be the mediator of neoangiogenesis and for tumor growth, and metastasis [110]. CXCL7, is the most abundant platelet chemokine, which is released from the activated platelets in the form of an inactive precursor known as connective tissue activating peptide III (CTAP-III), which undergoes proteolytic cleavage at the N-terminus resulting in formation of CXCL7/neutrophil activating peptide 2/(NAP2). CXCL7 interact with CXCR2 on neutrophils and down regulate the activity of CXCR2, and therefore works in a

negative feedback loop [111]. It has been observed that NAP2 degrades heparin and heparan sulfate which indicates its role in breakdown of basement membrane during metastasis, angiogenesis, and arthritis [111]. CXCL8 is involved in chemotaxis of T-cells, basophils, and cytokine stimulated eosinophils [10]. CXCL8 interacts with both CXCR1 and CXCR2 receptors thereby plays regulatory role in various signaling pathways. CXCL8 is involved in angiogenesis of endothelial cells, and also plays a vital role in migration of endothelial cells, neutrophils, and cancer cells at the tumor site. Hence, designing CXCL8 based inhibitors can be used to overcome various cancers [112].

1.14 Structural and oligomerization properties of NACs

NACs exist in reversible equilibrium of monomers and dimers. Some members of NACs can also form higher order oligomers including tetramers. NACs being the members of CXC chemokines, forms CXC type dimeric structures. Structures of all the NACs except CXCL3 and CXCL6, have been solved and characterized by either X-ray crystallography or NMR spectroscopy or both **(Table 1)**. Among CXC chemokines, CXCL8 was the first chemokine for which structure was solved by both X-ray crystallography [113] and NMR spectroscopy [114]. Tetramerization was observed in the crystal structure of CXCL7. Tetramer structure of CXCL7 is the resultant of back to back arrangement of two CXC type dimers of CXCL7 [48].

NACs, in addition to the formation of homodimers also undergo the process of hetero-dimerization, which further adds diversity to the mechanisms employed by chemokine systems to modulate their functional activities. Formation of heterodimer between the angiogenic chemokine CXCL8 and anti angiogenic chemokine CXCL4 has been reported. It was observed that the formation of CXCL4/CXCL8 hetero-dimer resulted an increase in anti-proliferative effect of CXCL4 on endothelial cells and an increase in the CXCL8 mediated migration of cells [115]. Molecular dynamic simulations were employed in order to determine the association free energies for formation of heterodimers CXC chemokines using CXCL1, CXCL7 CXCL8 and CXCL4. It is observed that some of pairs of chemokines are more favorable over others [116]. Recently, molecular basis of heterophilic interactions of CXCL7 with CXCL1, CXCL4, and CXCL8 chemokines have also been investigated [117,118]. It has been demonstrated that the favorable packing and ionic interactions resulted in the formation of stable heterodimers CXCL7-CXCL1 and CXCL7-CXCL4, but the presence of repulsive ionic interactions disfavored the formation of

CXCL7-CXCL8 heterodimer. Heterodimer interactions with GAG and chemokine receptor have also been characterized using CXCL7-CXCL1 trapped heterodimer [117]. It was observed that the heterodimer showed the same activity as native proteins for CXCR2 function in Ca^{2+} release assay but showed differential interactions with heparin as compared to CXCL7 monomer [117].

Table 1: *List of neutrophil activating chemokines (NACs) from human and mouse, along with PDB IDs, for which the structures are available.*

Name	Common Name	PDB ID	Method	Structure	Reference
HCXCL1	GRO α, MGSA	1MGS	NMR	Dimer	[119]
		1MSG	NMR	Dimer	[120]
		1MSH	NMR	Dimer	[120]
HCXCL2	GRO β	1NK	NMR	Dimer	[121]
MCXCL2	MIP2	1MI2	NMR	Dimer	[122]
MCXCL2	MIP2	3N52	X-Ray	Dimer	[123]
HCXCL5	ENA-78	2MGS	NMR	Dimer	[124]
HCXCL7	NAP2	1NAP	X-Ray	Tetramer	[48]
		------	NMR	Monomer	[125]
		1TVX	X-Ray	Tetramer	[126]
HCXCL8	IL8	1IL8	NMR	Dimer	[114]
		1ICW	X-Ray	Dimer	[127]
		1QE6	X-Ray	Dimer	[128]
		1ILQ	NMR	Dimer	[129]
		3IL8	X-Ray	Dimer	[130]
		1ILP	NMR	Dimer	[129]

1.15 Interaction of NACs with glycosaminoglycans

Numerous studies have been carried out to decipher the interaction of NACs with GAGs. GAG binding sites were defined on CXCL8 using trisulfated heparin disaccharide. The GAG binding surface includes the C-terminal helix and proximal 3_{10} helix residues (18-20) in CXCL8 [77]. A more recent study characterized the binding of monomer and dimer variants of IL-8 with GAGs of different chain lengths and showed that the pattern and extent of binding varies between monomer and dimer for different GAGs [131]. Various MD based and other docking studies have also been carried out for human CXCL8, which reported different binding modes of GAGs. In model I, GAGs binds parallel to the helices by their sulfated domains and oriented by their non-sulfated domains, this model is also known as horse shoe model [132,133]. In model II, GAGs

bind perpendicular to the helices spanning the dimer interface [134,135]. In model III, GAG binds parallel but lies in middle of two helices [135] **(Fig. 1.12)**. GAGs can bind to the monomers using either model I or model II as shown in **Fig. 1.12**.

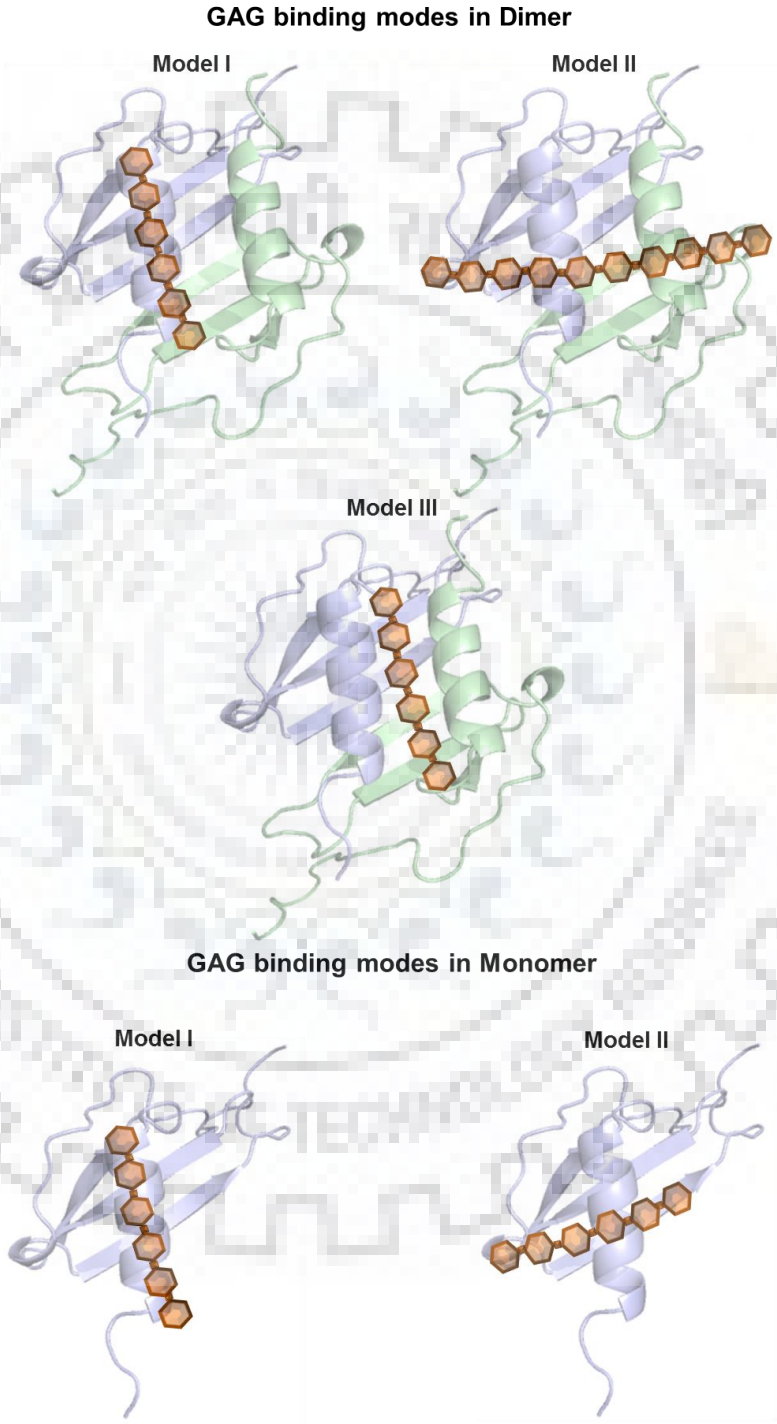


Figure 1.12: *Schematic showing the different models for GAG binding to the both monomeric and dimeric form of chemokines.*

Binding studies of murine CXCL1 with heparin octasaccharide revealed the perpendicular binding of GAGs and also figured out that the residues of 3_{10} helix (H20 and K22) and C-terminal helix (K62 and K66) serving as hotspots for GAG binding [136]. Geometry of murine CXCL2 with heparin disaccharide was also defined. They found the two major regions of GAG binding, one is 3_{10} helix (D19 and K21) and the other is C-terminal α helix (K61, K65 and K69) [123]. Binding of CXCL7 monomer to the heparin octasaccharide was mapped, and found that the heparin can bind in different modes and existence of structural plasticity in GAG binding interface [137]. This indicates the specificity of GAG binding geometry for chemokine surfaces [137]. Binding of CXCL5 to GAGs of different lengths ranging from dp4, dp8 and dp14 evidenced higher affinity for both dp8 and dp14 as compared to dp4. Stoichiometry for their binding is two GAG chains per CXCL5 [138].

Wealth of functional and biological information is available for chemokines. However, to understand their functions precisely, there is a need to correlate it with evolutionary profiles, structural properties, and their molecular interactions at atomic level. To discern these molecular features, various computational and biophysical techniques are available. In the following sections, some of the techniques used in the present study are described in detail.

1.16 Computational techniques

Computational biology is vast field that relates computational techniques with biology in order to decipher biological mechanisms. Bioinformatics is a multidisciplinary approach to analyze large biological data to get useful biological information. It deciphers the evolutionary aspects, decodes the expected protein products, determines the interactions of proteins with other biological macromolecules and generates structural models for new proteins, rationally designs novel drugs with improved binding properties. Considering the scope of present thesis, the computational techniques used to study the molecular evolution and protein structure modeling have been described in the following section.

1.16.1 Molecular evolution of proteins

Numerous bioinformatics tools have been developed to study the evolution of proteins to characterize evolution-structure-function relationships. They include multiple sequence alignment, substitution rates, phylogenetic analysis, selection criteria, and co-evolutionary analysis etc.

Multiple sequence alignment

This amino acid or nucleotide sequence alignment depicts the regions with high similarity and differences, that can be used to infer the residues that are highly conserved among the species, and contributes to the important function of the protein. On the other hand, the sites of nucleotide or amino acid differences can be implicated as positions responsible for the evolution of the gene or protein. The study of protein evolution begins with the search for the gene sequences from the different species from the large databases like GenBANK, EMBL, and DDBJ etc. This can be done by using BLASTn or BLASTp, where n and p represent nucleotide and protein respectively. BLAST is the basic local alignment search tool that works basically by searching for the sequences with local similarity. The sequences of a gene from different species are known as homologous sequences and more precisely the orthologous sequences. Once the orthologous sequences have been obtained, the next step is the alignment of those sequences using multiple sequence alignment (MSA) method. MSA aims at the placement of similar/identical alphabets, either nucleotides or amino acids from different sequences on the top of each other. Several MSA methods with high accuracy and speed are available, including Clustal W, T-Coffee, MUSCLE etc. [139].

Substitution rates

Once the multiple sequence alignment has been obtained, the alignment can be used to calculate the substitution rates or the evolutionary distances between all pairs of sequences. The evolutionary distance is the number of substitutions per site that have occurred between the sequences since they have originated from their common ancestor. MEGA, RevBayes, SATé-II, TransPhylo are some of the programs used for phylogeny/substitution [140-143]. Among them, MEGA (Molecular Evolutionary Genetics Analysis) is most popular. MEGA6.0, the latest version of MEGA provides number of models to determine the evolutionary distances among the sequences. Among them, p-distance method, computes the nucleotides proportions (p) at which the two sequences being compared are distinct. p-distance is calculated by dividing the number of nucleotide differences by the total number of nucleotides being compared. MEGA also yields the other related quantities such as transitions and transversions [140].

Phylogenetic tree

The phylogenetic tree also known as evolutionary tree is a branching diagram that shows the evolutionary relationship among the species. The branching pattern reveals the relatedness or the closeness of the species depending on their evolvement from the common ancestor. Closely related species exhibit a more recent common ancestor whereas the distantly related species exhibit a less common ancestor. Two types of phylogenetic trees can be constructed namely, the rooted tree and the unrooted tree. The rooted tree exhibits a root node that corresponds to the common ancestor to all the taxa. In contrast, the unrooted tree lacks the root node, but it won't hamper the information contained in the tree. Several methods have been developed for the construction of phylogenetic tree. They include: (1) distance based methods – UPGMA (unweighted pair group method with arithmetic mean), neighbor joining, and minimum evolution; (2) character based methods – maximum parsimony and maximum likelihood method. Reliability of the phylogenetic analysis is measured using bootstrap test. This test uses a resampling strategy and calculates the bootstrap value for each internal branch. This bootstrap value is in the percentage form, the higher percentage value reflects the more confidence for the correctness of the internal branch [144].

Selection analysis

Selection analysis is an essential criterion to decipher the evolutionary characteristics of the proteins at residue level. During the evolution of proteins, genes undergo several mutations that are either selected or lost. Changes that occur can be synonymous or non-synonymous. Synonymous changes are the neutral/silent changes in which a nucleotide substitution does not alter the protein sequence, whereas non-synonymous changes are amino acid replacements. The changes that occur are either positively selected or are negatively selected. Negative selection/purifying selection correspond to the prohibition of the spread of changes. The comparison of synonymous (dS) and non-synonymous substitution rates (dN) reveals the selection criteria followed for the given protein coding site. The difference between these two rates is calculated by the ratio omega (ω), which is given by dN/dS. The value of dN/dS >1 indicates the positive selection, dN/dS < 1 indicates the negative selection, dN/dS = 1 indicates the neutral selection. The Neutral selection has little effect on structure and function of protein. The dN/dS ratios can be calculated for each amino acid residue in the protein by using different programs including DATA MONKEY, Codeml (part of PAML package) etc. These programs use different codon based maximum likelihood methods to

identify the sites under positive and negative selection. Positively selected amino acid indicates that they have undergone changes to inculcate new functions to the protein. In contrast, the negatively selected amino acids represent the amino acids that are fixed and have not undergone any changes during the course of evolution and are important for the protein fold, stability and existing functionalities [145].

Concerted and divergent evolution

Different models of evolution have been defined for the evolution of multigene protein families including: (1) Concerted evolution, in which all the members of a gene family evolve in a concerted/coincidental manner. In concerted evolution, the mutation occurring in a repeat, will spread into the entire members of the gene family by repetition of unequal crossover or gene conversion, thus resulting in homogenization of DNA sequences of all the member genes. (2) Divergent evolution, in which the genes obtained after being duplicated from their parental genes, acquires new gene functions independently [146,147]. Evolution pattern followed by the genes can be predicted by the pair wise comparison of dN and dS values of different species. Higher dS values as compared to dN values indicates the concerted evolution of genes. In contrast, the higher dN values in comparison to dS correspond to divergent evolution [146,147].

Co-evolution analysis

The term co-evolution refers to the simultaneous evolution of pair of residues. Even though coevolving residues are present far away in the sequence, they end up in close proximity in the 3D-fold of the protein. Such a co-evolving pair provides the information about the correlated mutations/correlated amino acid substitutions prevailed in the protein family as a result of evolutionary changes. Lovell and Robertson defined the coevolution as "reciprocal evolutionary change at evolutionarily interacting loci" [148]. There are several synonyms for coevolution, which includes correlated mutation, co-variation, and co-substitution. MISTIC (Mutual Information Server to Infer Coevolution) is a web server that provides the platform to identify the pair of coevolving residues in the protein family based on their multiple sequence alignment [149].

1.16.2 Protein structure modeling

To date, structures for hundreds of proteins are available. However, still there is lag in the structural data for large number of proteins. For those proteins, considerable progress has been

made in developing protein 3D structure prediction methods. These methods have been divided into four classes: (1) homology modeling/template based method, (2) ab initio method (3) protein threading, (4) fragment based method [150].

Among these methods, the most facile and user friendly is the homology modeling, which is based on the extent of homology of the target sequence to that of reference protein. The basic assumption in the homology modeling is that the proteins having the same sequence will end in same structural fold [151]. More similarity/identity assures the accuracy of predicted structure. Sequences with > 40 % identity with template sequence are applied to predict the structures through homology modeling. Prediction of structure via homology modeling essentially requires the identification/selection of a template protein, whose PDB is present. Then the target sequence will be aligned with the template sequence and the modeling will be performed by using the structurally conserved regions (both backbone and side chain) of the template structure. The modeled structure will be assessed for its quality using structure assessment tools [150]. Numerous homology modeling methods/servers have been developed in the recent years. Some of the prominent servers are; SWISS-MODEL [152], MODELLER [153], META PP [154], ROBETTA [155].

1.17 Experimental tools to measure structure, stability, and molecular interactions of proteins

Biophysics and structural biology research is targeted towards understanding the concepts underlying the biological processes at molecular level. It aids in investigating the interactions between various biomolecules including nucleic acids, proteins, carbohydrates and their regulatory mechanisms. Biophysical characterization of proteins mainly involves the determination of their molecular sizes, their structural stability, folding/unfolding kinetics, behavior in different solution conditions, interactions with binding partners etc. Numerous biophysical techniques have been emerged for the biophysical and structural characterization of proteins. They include chromatography (affinity, size exclusion, ion exchange), spectroscopy techniques (UV/Vis, Circular Dichroism (CD), Fluorescence, NMR- Nuclear magnetic resonance, EPR- Electron paramagnetic resonance), microscopy (TEM -Transmission electron microscopy, SEM-Scanning electron microscopy , Confocal), SPR - surface plasmon resonance, X-ray crystallography, etc. The techniques employed in the present study have been described below.

1.17.1 Size exclusion chromatography

Size exclusion chromatography (gel filtration chromatography), is used to separate the molecules on the basis of their size **(Fig. 1.13)**. It is being widely used for the purification of biological polymers including proteins, polysaccharides, and nucleic acids etc. The basic principle behind this technique is the different permeation rates of particles of different sizes through the stationary phase [156]. Size exclusion chromatography is also being widely used to determine the oligomeric states of proteins [157].

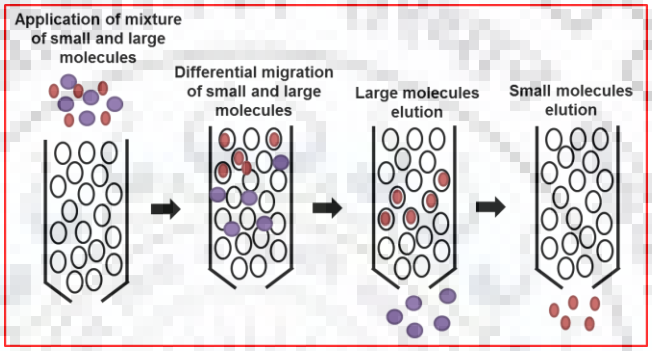


Figure 1.13: *Schematic illustrating the principle of size exclusion chromatography.*

1.17.2 Ion exchange chromatography

Ion exchange chromatography is used to separate the ionizable molecules on the basis of charge. The molecules of the similar sizes that cannot be separated by size exclusion chromatography can be separated by ion exchange chromatography based on the differential charges at the given pH conditions [158]. The ion exchange chromatography has been divided into two types, (a) cation exchange chromatography and (b) anion exchange chromatography based on the charge of their matrix. The proteins are zwitter ionic in nature due to the presence of both positively and negatively charged groups. Depending on the different pH conditions, proteins possess positive charge or negative charge or no charge. When the mixture of the protein is employed in the ion exchange column, the protein will bind to the column and the other undesired proteins and impurities will be washed out of the column, and then the protein can be eluted from the column either by changing the pH or using salt concentration gradient. The charged entities of the elution buffer will displace the bound protein charged groups. At lower salt concentrations, the proteins with small number of charged groups will be displaced by the buffer, and at higher

concentrations, the desired protein with maximum charged groups will be displaced by the salt charged groups, hence results in the elution of the desired protein.

Heparin binding assay

This assay is based on the cation exchange chromatography as the heparin is negatively charged polysaccharide. It is used to purify a wide variety of proteins including coagulation factors, DNA binding proteins, lipoproteins, enzymes, receptors etc. Heparin binds to the diversified proteins by the two mechanisms, one because of the affinity for the protein and other is by the ion exchange mechanism. Therefore, the heparin columns are being used to purify the proteins and also to assess their binding affinities.

Heparin binding assay is being widely used to assess the differential affinity of the heparin binding proteins and to determine the importance of particular amino acid residues for the GAG/heparin binding. Several studies have exploited this assay to decipher the GAG binding propensities for several of the proteins including chemokines, and determined the relevance of particular amino acid for GAG binding by generating the GAG binding incompetent mutants [159,160].

1.17.3 Circular dichroism (CD) spectroscopy

CD is the spectroscopic technique which is of great importance for understanding the structure, and folding of different proteins [161,162]. CD is the resultant of differential absorption of left and right handed circularly polarized light. When the plane polarized light passes through an asymmetric sample, the two circular components are absorbed to different extents **(Fig. 1.14)**.

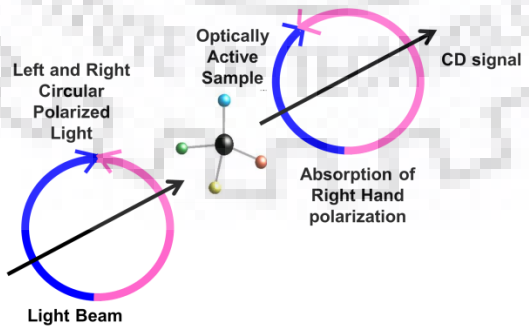


Figure 1.14: *Schematic demonstrating the principle of circular dichroism.*

Protein CD has been divided into two parts; one is far UV CD and other is near UV CD. Far UV CD also known as backbone CD, which is recorded in the range of 190 to 250 nm. It gives the basic structural information about the protein. The chromophore of far UV-CD is peptide bond. This implies that the peptide bonds present in α-helix, β-sheets, and random coil gives characteristic far UV-CD spectra **(Fig. 1.15)**. Thus, the analysis of such CD spectra can be used to calculate the fraction of each secondary structure when compared with reference protein spectrum.

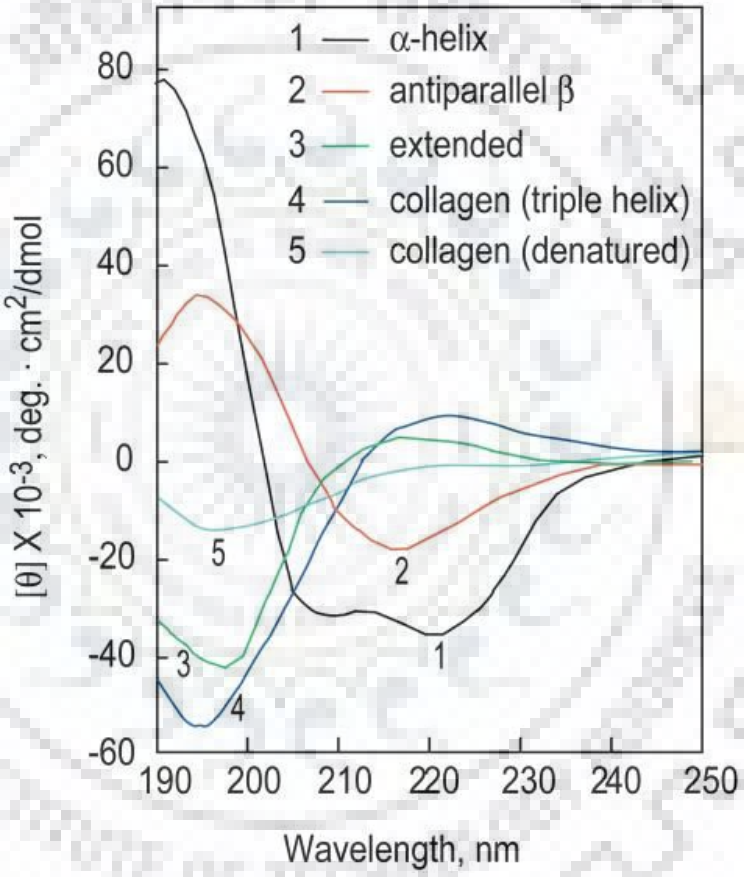


Figure 1.15: *CD spectra of poly-L-lysine in different conformations: α-helical, antiparallel β sheet, and extended conformation shown in black, red, and green respectively. Placental collagen in its native triple-helical and denatured form is shown in blue and cyan respectively. Figure adapted from Greenfield 2006 [162]. Reprinted by permission from Macmillan Publishers Ltd: [Nature Protocols] [162], copyright (2006).*

Near UV CD also known as aromatic CD, is recorded in the range of 250 to 300 nm. Aromatic amino acids doesn't exhibit any intrinsic CD signal but they exhibit an induced CD band

due to their presence in the chiral environment. Thus, this spectrum is sensitive to the tertiary structure of proteins in which the achiral aromatic side chains are present.

CD is measured in terms of ellipticity (ΔE). It is equal to the difference in absorption of E_R and E_L by achiral/asymmetric molecule. It can also be measured in terms of degrees of ellipticity which is given by an angle θ, whose tangent is the ratio of minor to the major axis. [θ], denotes the molar ellipticity and is given by the equation 1.1 given below

$$[\theta] = \theta/10 * C * l, \qquad \textbf{(1.1)}$$

Where, θ is ellipticity in mdegree, C is the molar concentration, and l is the path length of the cuvette in cm.

1.17.4 Fluorescence spectroscopy

Fluorescence is an emission phenomenon in which light absorbed at lower wavelengths is emitted at higher wavelengths. Fluorescence spectroscopy is a very important analytical tool for investigating structures, conformational stabilities, binding, and dynamics of biological systems with quantitative and qualitative information [163].

Basic principle of fluorescence is based on the fact that when the molecules absorb light of appropriate energy or wavelength, the electrons get excited to the higher energy levels. The excited state electrons come back to the ground state through number of radiative or non-radiative processes. Fluorescence belongs to one of those radiative processes as depicted in jablonski diagram **(Fig. 1.16)**.

Aromatic amino acids including tryptophan, tyrosine, and phenylalanine are the intrinsic fluorophores of proteins which if present in the protein makes the protein fluoresce. If the protein is deficient of these intrinsic fluorophores, extrinsic fluorophores such as 1-anilino-8-naphthalene sulfonate (ANS) can be used to delineate conformational and stability changes in the protein.

Fluorescence of molecules can be characterized by number of parameters that includes fluorescence intensity, life time, polarization, and quantum yield. All these parameters can be derived from the fluorescence spectra of the molecules and can be used to decode the specific structural and dynamic information about the molecules. Fluorescence quantum yield and life time are among the most important parameters.

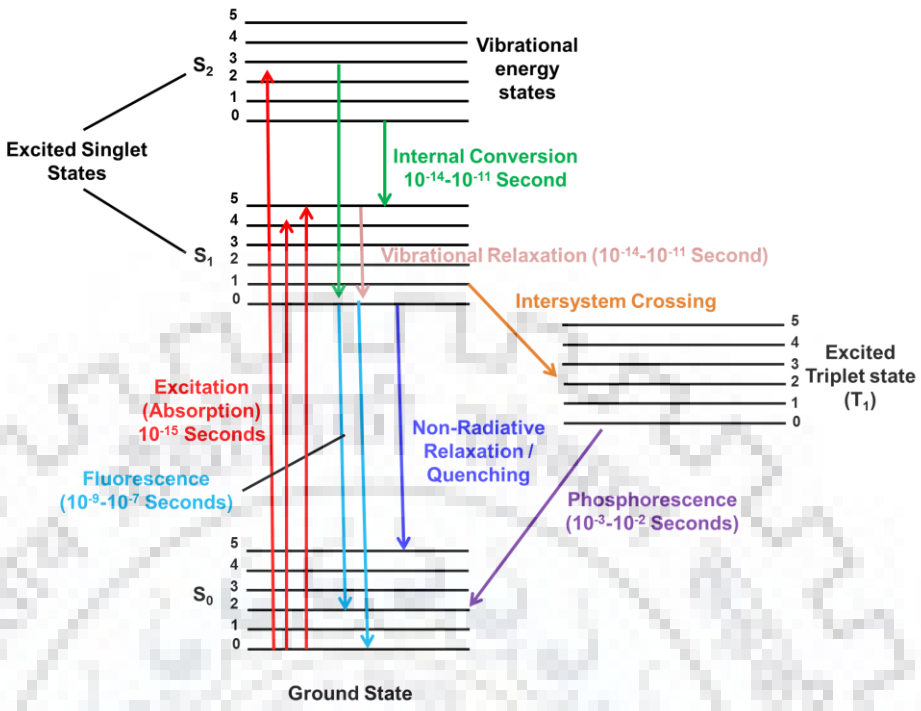


Figure 1.16: *Jablonski diagram showing the different physical processes that a molecule undergo after getting excited by the absorption phenomena that includes fluorescence, phosphorescence, internal conversion, inter-system crossing, vibrational relaxation, and collisional quenching. S and T denotes singlet and triplet state, respectively.*

Fluorescence quantum yield is the ratio of fluorescence photons emitted to the photons absorbed which is given by the following equation:

$$\phi_f = \frac{k_r}{k_r + k_{nr}} \qquad \textbf{(1.2)}$$

Where k_r denotes radiative rate constant and k_{nr} is non radiative rate constant.

Fluorescence life time (FLT) is an intrinsic property of the fluorophore and is defined as the average time the molecule spends in the excited state before reaching to its ground state by emission of photon. Fluorescence life time is represented by τ, which is given by the following equation:

$$\tau = \frac{1}{k_r + k_{nr}} \qquad \textbf{(1.3)}$$

Similar to radioactive decay is known as fluorescence decay which relates the fluorescence intensity at time t is proportional to the number of excited fluorophores.

$$I(t) = I_0 \, exp\,(- t / \tau_f) \qquad \textbf{(1.4)}$$

According to the equation **(1.4)**, fluorescence life time can be calculated as time required for the fluorescence intensity at time t to drop by 1/e or 37% of its original fluorescence intensity at time t = 0 upon excitation with light pulse.

1.17.5 Nuclear magnetic resonance (NMR) spectroscopy

Advances in NMR technology including NMR hardware (cryoprobes, pulse field gradients, increased field strengths), new methodologies [164,165], and isotopic labeling techniques [166,167] opened up its new avenues for biological NMR. Biological NMR essentially deals with the structure-function relationships of the biomolecules and their complexes, concepts of magnetic resonance imaging and metabolic profiling of the body fluids, cells tissues etc. [168-171]. Considering the scope of the thesis in the current NMR section, the concepts regarding protein NMR are discussed in detail. Protein NMR comprises of deciphering the three dimensional structure of proteins, calculation of conformational changes, stability parameters, dynamics in different time scales at residue level for denatured, partially folded, misfolded, and native state proteins across the protein folding funnel. Protein NMR is also being widely used in studying the interactions of proteins with a diverse class of biomolecules including proteins, nucleic acids, carbohydrates etc. at the atomic level. A brief out line of NMR methods used for protein structure calculation, dynamics have been discussed in the following sections.

1.17.5.1 Protein structure determination using NMR spectroscopy

Basic steps involved in protein structure determination using NMR spectroscopy have been summarized in **fig. 1.17**.

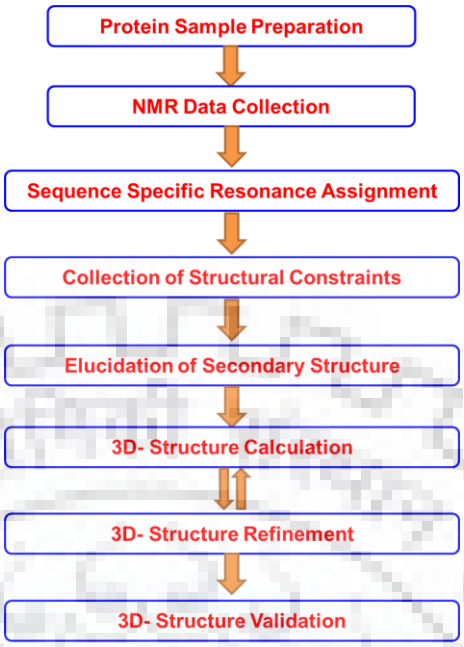


Figure 1.17: *Basic steps in involved in deciphering the protein structure using NMR spectroscopy.*

The foremost step in protein structure determination is to get the protein in hand, which involves the molecular biology work for expression and purification of labeled protein samples. Once the sample is prepared, the next step is to collect the NMR data by performing various NMR experiments. After the data has been collected, data needs to be analyzed which involves the chemical shift assignments for the backbone and side chains by a sequential walking along the backbone and side chains through dipolar correlations. Peaks obtained in 3D experiments will be correlated to specific amino acids by moving from plane to plane. Once assignment for all the amino acids of the polypeptide has been obtained, numerous structural parameters can be obtained from the data including coupling constants, inter-proton distances, and orientation of N-H and C-H bond vectors. These structural parameters are used to calculate structural ensembles based on distance geometry or by simulated annealing approach. Many programs have been designed for the process of generating the structure using NMR structural restraints including ARIA (Ambiguous Restraints for Iterative Assignment), CNS (Crystallography & NMR System), X-PLOR (exploration of conformational space of macromolecules) [172], DYANA (Dynamics algorithm for NMR applications) [173], CYANA (Combined assignment and dynamics algorithm for NMR applications) [174].

Isotopic labeling techniques

To overcome the problem of increased NMR spectral complexity for larger proteins and protein complexes, several labeling techniques have also been introduced including: (1) **Deuteration of specific protons** to reduce their contribution to transverse relaxation of ^{13}C spin attached to them [175]. (2) **Segmental isotope labeling** which reduces the spectral complexity based on examination of only specific labeled segments of protein by NMR [176-178]. Segmental isotopic labeling techniques include; (a) **trans splicing** method, (b) **expressed protein ligation** [179]. Segmental labeling method has not only been employed for the resonance assignment of large proteins but also used to determine the inter-domain interactions, their relative orientations and to monitor conformational changes induced by ligand binding, structures determination of glycoproteins and for studying protein- protein complexes [180]. (3) **Amino acid type ^{13}C ^{15}N selective labeling** in which the specific ^{13}C, ^{15}N labeled amino acid is used against an unlabeled ^{12}C, ^{14}N background [181,182]. (4) **Amino acid selective unlabeling or reverse labeling**, in which the cells are supplemented with an unlabeled selective amino acid against the ^{13}C, ^{15}N labeled background [183,184]. (5) **Selective methyl group protonation** involves the incorporation of a protonated chemical compound in the medium which acts as a precursor for Ala, Val, Leu, and Ile. This results in the ^{2}H, ^{13}C, ^{15}N-uniform labeling of the protein with protonation at the methyl groups of Ala, Val, Leu and γ^2 methyl site of Ile [185,186]. Other labeling methods include **stereospecific ^{2}H labeling** and **stereo array isotope labeling** which employs the use of stereospecific and regiospecific deuterium labeled amino acids (SAIL amino acids) [187,188].

Fast NMR methods

NMR has made tremendous progress both at technological and methodological levels. Efforts are being made in the NMR research with an aims to increase the speed for data collection and analysis. The increase in dimensionality of NMR leads to an increased acquisition time. This is due to an exponential increase in number of data points (that has to be collected in the indirect dimension). Hence, new methods have been developed for the rapid data collection for the multidimensional NMR spectra. These methods have been classified into six different types: (1) **spatial frequency encoding approach**, (2) **time sharing (simultaneous data acquisition) approach**, (3) **sparse sampling**, (4) **reduced dimensionality methods (projection NMR spectroscopy)**, (5) **Hadamard spectroscopy** and (6) **fast pulsing methods** [189].

Structures for large proteins

Novel NMR methods have been developed for high molecular weight proteins that include TROSY **(transverse relaxation optimized spectroscopy)**, which is based on the constructive interference of the dipole-dipole coupling and chemical shift anisotropy relaxation. Other method developed is CRINEPT **(correlated relaxation-enhanced polarization transfer)** that amalgamates INEPT **(insensitive nuclei enhanced by polarization transfer)** with CRIPT (**cross-correlated relaxation-induced polarization transfer)**. As when TROSY is applied for larger molecules during evolution and detection times, the INEPT transfers during evolution in those cases becomes limiting. Hence, CRINEPT which overcomes this limitation is used, as CRIPT is directly related to the size of the molecule. It results in efficient transfer of magnetization thus provides enhanced sensitivity in addition to that from TROSY. Several other methods are also being used for handling large proteins including Residual dipolar couplings (RDC) and Paramagnetic Relaxation Enhancement (PRE), Spectral perturbations [190-192] and Cross saturation [193,194] that yields information about the long range distance constraints, orientation of molecular domains, binding surfaces etc.

1.17.5.2 ***Structural elucidation of biomolecular complexes using NMR spectroscopy***

Besides, the involvement of NMR spectroscopy as an important tool for the retrieval of protein structures, NMR has also made tremendous contribution in determining the structures of bio-molecular complexes like protein:protein, protein:RNA, protein:DNA, protein:carbohydrates, oligomeric form of proteins, proteins in complex with membrane lipids, protein:chaperone, disordered proteins etc. [195-204]. Several NMR strategies either alone or in combination are being used to characterize the structure of the bio-molecular complexes. NOE derived inter-proton distances provides the valuable information to define the structure of macromolecular complex. Selective methyl group labeling techniques that aids in efficient methyl-methyl NOE measurements and the employment of methyl TROSY effects in NOESY experiments further eases the process of structure determination of macromolecular complexes [185,205-209].

In addition to the structure determination of complexes, NMR also aids in the determination of interactive sites on the individual components of protein complexes.

One simplest method is ***chemical shift mapping*** of chemical shift perturbation (CSP). Chemical shift of NMR active nuclei arises as a result of electronic environment in which the

nuclei are present. Any change in the electronic environment of the nucleus results in change in the chemical shift of the nucleus. Binding events or the interactions among the macromolecules leads to the changes in the chemical shift of the interacting nuclei. Such changes can define the specific residues participating in the interaction between the interacting partners. Analysis of such spectral shifts/perturbations in a site specific manner is known as **chemical shift mapping/chemical shift perturbation** [191,192,210].

Numerous other methods are also available to map these interactions includes: **NOEs** (Nuclear Overhauser effect), which provides the evidence for the spatial proximation of the macromolecules in the complexes. **Isotopic filtered NOE** methods are being employed for the determination of intermolecular NOEs in differentially labeled complexes [211]. A modified version of isotopic filtered NOE known as **REDuced/Standard PRoton density INTerface (REDSPRINT)** is also being used to study intermolecular interfaces [212]. Other less labour intensive cross relaxation method known as **saturation transfer/cross saturation**, in which the interfacial residues are determined by observing the changes in the peak intensities in 2D HSQC from one of the binding partner when the signals from the partner are saturated using radiofrequency (RF) pulse [213,214].

1.17.5.3 NMR parameters for structural and dynamic characterization of proteins

Amide proton temperature coefficients

Amide proton chemical shifts are highly sensitive to the environment in which the amide protons are present. Moreover amide protons involve in either intermolecular or intramolecular hydrogen bonding. In general, there should be a linear dependence of amide proton chemical shift and temperature [215-219]. Therefore, an increase in temperature leads to a change in amide proton chemical shift, which can be directly correlated to the H-bonding and local stability of the protein. Thus, amide protons involved in strong H-bonding shows a lower gradient of change with temperature, as compared to those involved in weak or not involved in any hydrogen bonding. This indicates that gradient of chemical shift change which is known as temperature coefficient, can be used to determine the strength of hydrogen bonding, in which the amide protons are involved. Temperature coefficient value ranges from - 2 to - 4 ppb/K for amide protons that are involved in strong hydrogen bonding in proteins. Temperature dependence experiments were carried out for Lysozyme and BPTI (basic pancreatic trypsin inhibitor) and a linear pattern of amide proton

chemical shifts was observed with temperature in the range of 278 – 328 K for Lysozyme and 279-359 K for BPTI [215]. The temperature coefficients were found to be in the range of -16 to + 4 ppb/K. The values were more positive than -4.5 ppb/K for strongly hydrogen bonded amide protons [215]. This is due to more lengthening of intermolecular hydrogen bond than the intramolecular hydrogen bonds. This is the case with the rigid proteins, but if the proteins are dynamic as is the actual case, then the amide proton chemical shifts will also be affected by the structural and dynamic changes and their temperature dependence will be no longer linear. The nonlinear temperature dependence of amide proton chemical shifts indicates that amino acid is accessible to multiple conformational states [217,218]. Elucidation of such amino acids in proteins gives an overview of its structural dynamics.

Hydrogen exchange

Backbone amide hydrogen-deuterium exchange is another unique NMR experiment to characterize the residue level structural stability of proteins [220-223]. The phenomenon of hydrogen deuterium exchange is based on the hydrogen exchange reaction that occurs when an amide proton exchanges with another solvent proton/deuteron (H-D exchange). The exchange of amide protons with deuterons depends on the accessibility of residues to the solvent. Residues present in secondary structures and hydrophobic core/protein core will not be accessible to the solvent because of their protection by hydrogen bonding and hence will not undergo H/D exchange easily. The residues accessible to the bulk solvent (D2O) will undergo the H/D exchange process rapidly. The exchange rates thus provide a clear picture about the solvent accessibilities of amide protons. Protons not involved in hydrogen bonding will possess high exchange rates as compared to those involved in hydrogen bonding or in hydrophobic core of the protein. Therefore, hydrogen deuterium exchange studies are being widely used to study structural stabilities, dynamics, inter molecular interactions, and folding landscape of proteins.

NMR relaxation measurements

Proteins undergo various conformational changes while carrying out their specific functions [224]. Flexibility in the protein structures also contributes to their adaptability to perform multiple functions and also to evolve the novel functions [225]. Solution state NMR provides us an opportunity to characterize the residue level dynamic behavior of proteins at different time scales. Relaxation is used to monitor the global and local motions in the proteins, by observing the

restoration of spins to equilibrium state that has been perturbed previously. Heteronuclei ^{15}N and ^{13}C are used for relaxation studies as they provide the isolated spin systems for which the relaxations phenomena are easily identifiable. Various types of relaxation parameters include longitudinal relaxation/spin-lattice relaxation (T_1), transverse relaxation/spin-spin relaxation (T_2) and steady state het NOE.

These NMR relaxation parameters are highly sensitivity to probe motions in proteins on picosecond to millisecond time scale. Thus, are being widely employed for characterizing the backbone dynamics and overall molecular tumbling motions in proteins [226,227]. Heteronuclear steady state NOE probes very high frequency motions in the protein backbone. Negative NOEs indicate the presence of large amplitude motions on sub nanosecond time scale, which are characteristic to unfolded protein. Such negative NOE values are also observed in folded proteins in the more flexible regions of loops and at the N-/C- termini. R_2 relaxation is sensitive to slow motions that are occurring in the range of milli to micro second time scales. Higher R_2 values indicate the presence of different conformational states in the protein. New strategies are being employed for further fine tuning the concepts of dynamics in the proteins. Transverse relaxation experiments including CPMG compensated relaxation dispersion have been developed which are useful in quantifying the different conformational states of the protein [228]. These relaxation parameters contributes a lot in understanding the sequence dependent motional restrictions and flexibilities both in native state and denatured state of proteins, and hence aided in tracking the dynamically driven protein folding and functionalities [229,230].

Translational diffusion NMR spectroscopy

In solution, molecules exhibit translational motions, also known as Brownian motions or self-diffusion. Self-diffusion of molecules depends on various physical factors including, shape, size, solvent viscosity, and other conditions like temperature, pH etc. [231]. Diffusion coefficient of the molecule (D) is calculated by the Stokes-Einstein relation (assuming the spherical shape of the molecule):

$$D = k\,T/6\,\pi\,\eta\,r_s \qquad \textbf{(1.5)}$$

Where, K is the Boltzmann constant, T is temperature, η is the viscosity of solvent, and r_s is the hydrodynamic radius of the molecule.

The translational information of the molecule can be encoded in the NMR spectrum by using pulse field gradient (PFG) NMR experiments, commonly known as diffusion ordered NMR spectroscopy (DOSY). 2D-DOSY spectrum presents chemical shift on the one dimension and diffusion coefficient constants on other dimension. The relation between the diffusion coefficient and molecular size is used to determine the molecular size, and also allows the identification of different molecules based on their sizes. DOSY can also be used to determine the solution properties of proteins including oligomerization and aggregation states, interaction of proteins with other macromolecules, differential conformations etc.

1.18 Thesis outline and objectives

The main objective of the thesis is to unravel the molecular aspects of a sub group of neutrophil activating chemokines known as GRO Chemokines.

GRO chemokines comprise of three highly identical NACs known as GRO α (CXCL1), GRO β (CXCL2) and GRO γ (CXCL3) that are expressed during melanoma tumors. These chemokines have been shown to be continuously expressed in melanoma tumors therefore, involves in growth and progression of tumors [232]. GRO gene was first reported from its constitutive expression in a highly tumorigenic Chinese hamster cell [233]. GRO α/MGSA (Melanoma growth stimulating activating) protein was purified and characterized that was able to stimulate the growth of melanoma cells, found identical to GRO [234]. Two other closely related GRO genes were then identified, characterized, and were named as GRO β and GRO γ [235]. They observed differential expression patterns of all three GRO genes, regulated in tissue and signal specific manner. GRO family chemokines acts as arrest chemokines that interact with CXCR2 and results in monocyte adhesion on vascular cell adhesion molecule (VCAM)-1 under flow in the presence of P-selectin [236]. Despite their close relativity, they are involved in different functions, activation, and expression patterns or pathways during several pathological and physiological conditions [237-241]. It has also been proposed that GRO α can act as a novel diagnostic marker for age-related pathology, including cancer [242]. Recently, GRO γ has been developed as a therapeutic agent for the treatment of medullablastoma [243].

The differential behaviors of these highly related members of GRO chemokines can be implicated as a result of their differential evolution patterns, fine adjustments in their structure, and dynamics, contributing to their differential avidities for oligomerization (homo/hetero) and their interactions with cellular partners including GAGs and receptors. Considering the biological importance of GRO chemokines and lack of molecular knowledge about their biophysical characteristics, the thesis has been designed to unravel their molecular interactions. In detail, the specific objectives of the thesis are as follows

1.18.1 Specific objectives

- Mechanistic insights into molecular evolution of species specific differential glycosaminoglycan binding surfaces in GRO Chemokines.
- Deciphering the oligomerization and GAG binding features of CXCL1/CXCL2 chemokines.
- Biophysical characterization of CXCL3 and its comparison with CXCL2.

1.18.2 Scope of the thesis

Considering the above specific objectives, **chapters 2-4** were designed as follows: **Chapter 2** unraveled the evolutionary perspectives of GRO chemokines using various computational tools and integrated the evolutionary findings with structure and function of GRO chemokines with special emphasis on GAG binding. **Chapter 3** deciphered the homo and hetero-oligomerization potentials of GRO chemokines (GRO α and GRO β) for which the structures are available. In addition, the effect of different GAGs/GAG mimetics on homo-/hetero-oligomerization of GRO α and GRO β were elucidated. **Chapter 4** unraveled the biophysical and structural characteristics of GRO γ, and their comparison with other GRO chemokines.

1.19 References

1. Newton K, Dixit VM. Signaling in innate immunity and inflammation. Cold Spring Harb. Perspect. Biol. 4(3) (2012).

2. Thomas J.Kindt and Barbara A.Osborne. Kuby Immunology, Chapter1: Overview of the immune system. (2006)

3. Ota T, Nei M. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol. Biol Evol. 11(3), 469-482 (1994).

4. Hughes AL, Nei M. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. Mol. Biol Evol. 7(6), 491-514 (1990).

5. Zhang J, Dyer KD, Rosenberg HF. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proc. Natl. Acad. Sci. U. S. A. 97(9), 4701-4706 (2000).

6. Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome Biol. 7(12), 243- (2006).

7. Roach DR, Bean AG, Demangel C, France MP, Briscoe H, Britton WJ. TNF regulates chemokine induction essential for cell recruitment, granuloma formation, and clearance of mycobacterial infection. J Immunol. 168(9), 4620-4627 (2002).

8. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. Biochim. Biophys. Acta. 1843(11), 2563-2582 (2014).

9. Borish LC, Steinke JW. 2. Cytokines and chemokines. J Allergy Clin. Immunol. 111(2 Suppl), S460-S475 (2003).

10. Steinke JW, Borish L. 3. Cytokines and chemokines. J Allergy Clin. Immunol. 117(2 Suppl Mini-Primer), S441-S445 (2006).

11. Zlotnik A, Yoshie O. The chemokine superfamily revisited. Immunity. 36(5), 705-716 (2012).

12. Widdison S, Coffey TJ. Cattle and chemokines: evidence for species-specific evolution of the bovine chemokine system. Anim Genet. 42(4), 341-353 (2011).

13. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ. Defining the origins and evolution of the chemokine/chemokine receptor system. J. Immunol. 176(1), 401-415 (2006).

14. Dubrac A, Quemener C, Lacazette E, Lopez F, Zanibellato C, Wu WG, Bikfalvi A, Prats H. Functional divergence between 2 chemokines is conferred by single amino acid change. Blood. 116(22), 4703-4711 (2010).

15. Yoshie O, Imai T, Nomiyama H. Chemokines in immunity. Adv. Immunol. 78, 57-110 (2001).

16. Thornton JW, DeSalle R. Gene family evolution and homology: genomics meets phylogenetics. Annu. Rev. Genomics Hum. Genet. 1, 41-73 (2000).

17. Wagner A. Birth and death of duplicated genes in completely sequenced eukaryotes. Trends Genet. 17(5), 237-239 (2001).

18. Bacon K, Baggiolini M, Broxmeyer H, Horuk R, Lindley I *et al.* Chemokine/chemokine receptor nomenclature. J. Interferon Cytokine Res. 22(10), 1067-1068 (2002).

19. Pease JE, Wang J, Ponath PD, Murphy PM. The N-terminal extracellular segments of the chemokine receptors CCR1 and CCR3 are determinants for MIP-1alpha and eotaxin binding, respectively, but a second domain is essential for efficient receptor activation. J. Biol. Chem. 273(32), 19972-19976 (1998).

20. Monteclaro FS, Charo IF. The amino-terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. Receptor activation by a pseudo-tethered ligand. J. Biol. Chem. 272(37), 23186-23190 (1997).

21. Mantovani A, Bonecchi R, Locati M. Tuning inflammation and immunity by chemokine sequestration: decoys and more. Nat. Rev. Immunol. 6(12), 907-918 (2006).

22. Ulvmar MH, Hub E, Rot A. Atypical chemokine receptors. Exp. Cell Res. 317(5), 556-568 (2011).

23. Mantovani A. The chemokine system: redundancy for robust outputs. Immunol. Today. 20(6), 254-257 (1999).

24. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, Rousseau S, Eidelman DH, Hamid Q. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. J Immunol. 191(5), 2731-2741 (2013).

25. Devalaraja MN, Richmond A. Multiple chemotactic factors: fine control or redundancy? Trends Pharmacol. Sci. 20(4), 151-156 (1999).

26. Baggiolini M. Chemokines and leukocyte traffic. Nature. 392(6676), 565-568 (1998).

27. Rot A, von Andrian UH. Chemokines in innate and adaptive host defense: basic chemokinese grammar for immune cells. Annu. Rev. Immunol. 22, 891-928 (2004).

28. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat. Rev. Immunol. 13(3), 159-175 (2013).

29. Richmond A, Yang J, Su Y. The good and the bad of chemokines/chemokine receptors in melanoma. Pigment Cell Melanoma Res. 22(2), 175-186 (2009).

30. Esche C, Stellato C, Beck LA. Chemokines: key players in innate and adaptive immunity. J. Invest Dermatol. 125(4), 615-628 (2005).

31. Sallusto F, Baggiolini M. Chemokines and leukocyte traffic. Nat. Immunol. 9(9), 949-952 (2008).

32. Luster AD. The role of chemokines in linking innate and adaptive immunity. Curr. Opin. Immunol. 14(1), 129-135 (2002).

33. Raman D, Sobolik-Delmaire T, Richmond A. Chemokines in health and disease. Exp. Cell Res. 317(5), 575-589 (2011).

34. Li M, Hale JS, Rich JN, Ransohoff RM, Lathia JD. Chemokine CXCL12 in neurodegenerative diseases: an SOS signal for stem cell-based repair. Trends Neurosci. 35(10), 619-628 (2012).

35. Christopherson KW, Hromas RA. Endothelial chemokines in autoimmune disease. Curr. Pharm. Des. 10(2), 145-154 (2004).

36. Tuttolomondo A, Di RD, Pecoraro R, Arnao V, Pinto A, Licata G. Atherosclerosis as an inflammatory disease. Curr. Pharm. Des. 18(28), 4266-4288 (2012).

37. Gerard C, Rollins BJ. Chemokines and disease. Nat. Immunol. 2(2), 108-115 (2001).

38. O'Hayre M, Salanga CL, Handel TM, Allen SJ. Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment. Biochem. J. 409(3), 635-649 (2008).

39. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. Cold Spring Harb. Perspect. Med. 2(8) (2012).

40. Wilen CB, Tilton JC, Doms RW. Molecular mechanisms of HIV entry. Adv. Exp. Med. Biol. 726, 223-242 (2012).

41. Schall TJ, Proudfoot AE. Overcoming hurdles in developing successful drugs targeting chemokine receptors. Nat. Rev. Immunol. 11(5), 355-363 (2011).

42. Fernandez EJ, Lolis E. Structure, function, and inhibition of chemokines. Annu. Rev. Pharmacol. Toxicol. 42, 469-499 (2002).

43. Goodsell DS, Olson AJ. Structural symmetry and protein function. Annu. Rev. Biophys. Biomol. Struct. 29, 105-153 (2000).

44. Wang X, Sharp JS, Handel TM, Prestegard JH. Chemokine oligomerization in cell signaling and migration. Prog. Mol. Biol. Transl. Sci. 117, 531-578 (2013).

45. Ali MH, Imperiali B. Protein oligomerization: how and why. Bioorg. Med. Chem. 13(17), 5013-5020 (2005).

46. Jones BA, Beamer M, Ahmed S. Fractalkine/CX3CL1: a potential new target for inflammatory diseases. Mol. Interv. 10(5), 263-270 (2010).

47. Salanga CL, Handel TM. Chemokine oligomerization and interactions with receptors and glycosaminoglycans: the role of structural dynamics in function. Exp. Cell Res. 317(5), 590-601 (2011).

48. Malkowski MG, Wu JY, Lazar JB, Johnson PH, Edwards BF. The crystal structure of recombinant human neutrophil-activating peptide-2 (M6L) at 1.9-A resolution. J. Biol. Chem. 270(13), 7077-7087 (1995).

49. Murphy JW, Yuan H, Kong Y, Xiong Y, Lolis EJ. Heterologous quaternary structure of CXCL12 and its relationship to the CC chemokine family. Proteins. 78(5), 1331-1337 (2010).

50. Salanga CL, O'Hayre M, Handel T. Modulation of chemokine receptor activity through dimerization and crosstalk. Cell Mol. Life Sci. 66(8), 1370-1386 (2009).

51. Kramp BK, Sarabi A, Koenen RR, Weber C. Heterophilic chemokine receptor interactions in chemokine signaling and biology. Exp. Cell Res. 317(5), 655-663 (2011).

52. Nesmelova IV, Sham Y, Gao J, Mayo KH. CXC and CC chemokines form mixed heterodimers: association free energies from molecular dynamics simulations and experimental correlations. J Biol. Chem. 283(35), 24155-24166 (2008).

53. Guan E, Wang J, Norcross MA. Identification of human macrophage inflammatory proteins 1alpha and 1beta as a native secreted heterodimer. J Biol. Chem. 276(15), 12404-12409 (2001).

54. von HP, Koenen RR, Sack M, Mause SF, Adriaens W, Proudfoot AE, Hackeng TM, Weber C. Heterophilic interactions of platelet factor 4 and RANTES promote monocyte arrest on endothelium. Blood. 105(3), 924-930 (2005).

55. Proudfoot AE, Power CA, Rommel C, Wells TN. Strategies for chemokine antagonists as therapeutics. Semin. Immunol. 15(1), 57-65 (2003).

56. Handel TM, Johnson Z, Crown SE, Lau EK, Proudfoot AE. Regulation of protein function by glycosaminoglycans--as exemplified by chemokines. Annu. Rev. Biochem. 74, 385-410 (2005).

57. Gulati K, Poluri KM. Mechanistic and therapeutic overview of glycosaminoglycans: the unsung heroes of biomolecular signaling. Glycoconj. J. 33(1), 1-17 (2016).

58. Gulati K, Meher MK, Poluri KM. Glycosaminoglycan-based resorbable polymer composites in tissue refurbishment. Regen. Med. 12(4), 431-457 (2017).

59. Hoogewerf AJ, Kuschert GS, Proudfoot AE, Borlat F, Clark-Lewis I, Power CA, Wells TN. Glycosaminoglycans mediate cell surface oligomerization of chemokines. Biochemistry. 36(44), 13570-13578 (1997).

60. Webb LM, Ehrengruber MU, Clark-Lewis I, Baggiolini M, Rot A. Binding to heparan sulfate or heparin enhances neutrophil responses to interleukin 8. Proc. Natl. Acad. Sci. U. S. A. 90(15), 7158-7162 (1993).

61. Cadene M, Boudier C, de Marcillac GD, Bieth JG. Influence of low molecular mass heparin on the kinetics of neutrophil elastase inhibition by mucus proteinase inhibitor. J. Biol. Chem. 270(22), 13204-13209 (1995).

62. Wagner L, Yang OO, Garcia-Zepeda EA, Ge Y, Kalams SA, Walker BD, Pasternack MS, Luster AD. Beta-chemokines are released from HIV-1-specific cytolytic T-cell granules complexed to proteoglycans. Nature. 391(6670), 908-911 (1998).

63. Ellyard JI, Simson L, Bezos A, Johnston K, Freeman C, Parish CR. Eotaxin selectively binds heparin. An interaction that protects eotaxin from proteolysis and potentiates chemotactic activity in vivo. J. Biol. Chem. 282(20), 15238-15247 (2007).

64. Lebel-Haziv Y, Meshel T, Soria G, Yeheskel A, Mamon E, Ben-Baruch A. Breast cancer: coordinated regulation of CCL2 secretion by intracellular glycosaminoglycans and chemokine motifs. Neoplasia. 16(9), 723-740 (2014).

65. Middleton J, Neil S, Wintle J, Clark-Lewis I, Moore H, Lam C, Auer M, Hub E, Rot A. Transcytosis and surface presentation of IL-8 by venular endothelial cells. Cell. 91(3), 385-395 (1997).

66. Roscic-Mrkic B, Fischer M, Leemann C, Manrique A, Gordon CJ, Moore JP, Proudfoot AE, Trkola A. RANTES (CCL5) uses the proteoglycan CD44 as an auxiliary receptor to mediate cellular activation signals and HIV-1 enhancement. Blood. 102(4), 1169-1177 (2003).

67. Wang L, Fuster M, Sriramarao P, Esko JD. Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. Nat. Immunol. 6(9), 902-910 (2005).

68. Schlorke D, Thomas L, Samsonov SA, Huster D, Arnhold J, Pichert A. The influence of glycosaminoglycans on IL-8-mediated functions of neutrophils. Carbohydr. Res. 356, 196-203 (2012).

69. Panitz N, Theisgen S, Samsonov SA, Gehrcke JP, Baumann L, Bellmann-Sickert K, Kohling S, Pisabarro MT, Rademann J, Huster D, Beck-Sickinger AG. The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. Glycobiology. 26(11), 1209-1221 (2016).

70. Nordsieck K, Pichert A, Samsonov SA, Thomas L, Berger C, Pisabarro MT, Huster D, Beck-Sickinger AG. Residue 75 of interleukin-8 is crucial for its interactions with glycosaminoglycans. Chembiochem. 13(17), 2558-2566 (2012).

71. Mobius K, Nordsieck K, Pichert A, Samsonov SA, Thomas L, Schiller J, Kalkhof S, Teresa PM, Beck-Sickinger AG, Huster D. Investigation of lysine side chain interactions of interleukin-8 with heparin and other glycosaminoglycans studied by a methylation-NMR approach. Glycobiology. 23(11), 1260-1269 (2013).

72. Kunze G, Kohling S, Vogel A, Rademann J, Huster D. Identification of the Glycosaminoglycan Binding Site of Interleukin-10 by NMR Spectroscopy. J Biol Chem. 291(6), 3100-3113 (2016).

73. Kunze G, Gehrcke JP, Pisabarro MT, Huster D. NMR characterization of the binding properties and conformation of glycosaminoglycans interacting with interleukin-10. Glycobiology. 24(11), 1036-1049 (2014).

74. Hofmann T, Samsonov SA, Pichert A, Lemmnitzer K, Schiller J, Huster D, Pisabarro MT, von BM, Kalkhof S. Structural analysis of the interleukin-8/glycosaminoglycan interactions by amide hydrogen/deuterium exchange mass spectrometry. Methods. 89, 45-53 (2015).

75. Kuschert GS, Coulin F, Power CA, Proudfoot AE, Hubbard RE, Hoogewerf AJ, Wells TN. Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. Biochemistry. 38(39), 12959-12968 (1999).

76. Pichert A, Samsonov SA, Theisgen S, Thomas L, Baumann L, Schiller J, Beck-Sickinger AG, Huster D, Pisabarro MT. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. Glycobiology. 22(1), 134-145 (2012).

77. Kuschert GS, Hoogewerf AJ, Proudfoot AE, Chung CW, Cooke RM, Hubbard RE, Wells TN, Sanderson PN. Identification of a glycosaminoglycan binding surface on human interleukin-8. Biochemistry. 37(32), 11193-11201 (1998).

78. Chakravarty L, Rogers L, Quach T, Breckenridge S, Kolattukudy PE. Lysine 58 and histidine 66 at the C-terminal alpha-helix of monocyte chemoattractant protein-1 are essential for glycosaminoglycan binding. J Biol Chem. 273(45), 29641-29647 (1998).

79. Lau EK, Paavola CD, Johnson Z, Gaudry JP, Geretti E, Borlat F, Kungl AJ, Proudfoot AE, Handel TM. Identification of the glycosaminoglycan binding site of the CC chemokine, MCP-1: implications for structure and function in vivo. J. Biol. Chem. 279(21), 22294-22305 (2004).

80. Hemmerich S, Paavola C, Bloom A, Bhakta S, Freedman R *et al.* Identification of residues in the monocyte chemotactic protein-1 that contact the MCP-1 receptor, CCR2. Biochemistry. 38(40), 13013-13025 (1999).

81. Johnson Z, Proudfoot AE, Handel TM. Interaction of chemokines and glycosaminoglycans: a new twist in the regulation of chemokine function with opportunities for therapeutic intervention. Cytokine Growth Factor Rev. 16(6), 625-636 (2005).

82. Proudfoot AE, Handel TM, Johnson Z, Lau EK, LiWang P, Clark-Lewis I, Borlat F, Wells TN, Kosco-Vilbois MH. Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. Proc. Natl. Acad. Sci. U. S. A. 100(4), 1885-1890 (2003).

83. Koopmann W, Krangel MS. Identification of a glycosaminoglycan-binding site in chemokine macrophage inflammatory protein-1alpha. J. Biol. Chem. 272(15), 10103-10109 (1997).

84. Koopmann W, Ediriwickrema C, Krangel MS. Structure and function of the glycosaminoglycan binding site of chemokine macrophage-inflammatory protein-1 beta. J. Immunol. 163(4), 2120-2127 (1999).

85. Proudfoot AE, Fritchley S, Borlat F, Shaw JP, Vilbois F, Zwahlen C, Trkola A, Marchant D, Clapham PR, Wells TN. The BBXB motif of RANTES is the principal site for heparin binding and controls receptor selectivity. J. Biol. Chem. 276(14), 10620-10626 (2001).

86. Amara A, Lorthioir O, Valenzuela A, Magerus A, Thelen M, Montes M, Virelizier JL, Delepierre M, Baleux F, Lortat-Jacob H, renzana-Seisdedos F. Stromal cell-derived factor-1alpha associates with heparan sulfates through the first beta-strand of the chemokine. J. Biol. Chem. 274(34), 23916-23925 (1999).

87. Campanella GS, Lee EM, Sun J, Luster AD. CXCR3 and heparin binding sites of the chemokine IP-10 (CXCL10). J. Biol. Chem. 278(19), 17066-17074 (2003).

88. Ingber DE, Mow VC, Butler D, Niklason L, Huard J, Mao J, Yannas I, Kaplan D, Vunjak-Novakovic G. Tissue engineering and developmental biology: going biomimetic. Tissue Eng. 12(12), 3265-3283 (2006).

89. Ko IK, Lee SJ, Atala A, Yoo JJ. In situ tissue regeneration through host stem cell recruitment. Exp. Mol. Med. 45, e57- (2013).

90. Spiess K, Jeppesen MG, Malmgaard-Clausen M, Krzywkowski K, Dulal K *et al.* Rationally designed chemokine-based toxin targeting the viral G protein-coupled receptor US28 potently inhibits cytomegalovirus infection in vivo. Proc. Natl. Acad. Sci. U. S. A. 112(27), 8427-8432 (2015).

91. Andreas K, Sittinger M, Ringe J. Toward in situ tissue engineering: chemokine-guided stem cell recruitment. Trends Biotechnol. 32(9), 483-492 (2014).

92. Badylak SF, Nerem RM. Progress in tissue engineering and regenerative medicine. Proc. Natl. Acad. Sci. U. S. A. 107(8), 3285-3286 (2010).

93. Gulati K and Poluri KM. Chemoattractants, Scaffolds and Endogenous Stem Cells: Adorable Partners of *In Situ* Tissue Regeneration. Austin J Biotechnol Bioeng.2(4), 1052- (2015)

94. Wang NX, Sieg SF, Lederman MM, Offord RE, Hartley O, von Recum HA. Using glycosaminoglycan/chemokine interactions for the long-term delivery of 5P12-RANTES in HIV prevention. Mol. Pharm. 10(10), 3564-3573 (2013).

95. von Recum HA and Hijaz A. Glycosaminoglycans for chemokine drug delivery.US20140364360 A1 (2014)

96. Purcell BP, Elser JA, Mu A, Margulies KB, Burdick JA. Synergistic effects of SDF-1alpha chemokine and hyaluronic acid release from degradable hydrogels on directing bone marrow derived cell homing to the myocardium. Biomaterials. 33(31), 7849-7857 (2012).

97. Pacheco DP, Reis RL, Correlo VM, Marques AP. The Crosstalk between Tissue Engineering and Pharmaceutical Biotechnology: Recent Advances and Future Directions. Curr. Pharm. Biotechnol. 16(11), 1012-1023 (2015).

98. Martins-Green M, Petreaca M, Wang L. Chemokines and Their Receptors Are Key Players in the Orchestra That Regulates Wound Healing. Adv. Wound. Care (New Rochelle.). 2(7), 327-347 (2013).

99. Jeyaseelan S, Manzer R, Young SK, Yamamoto M, Akira S, Mason RJ, Worthen GS. Induction of CXCL5 during inflammation in the rodent lung involves activation of alveolar epithelium. Am. J. Respir. Cell Mol. Biol. 32(6), 531-539 (2005).

100. Chavey C, Lazennec G, Lagarrigue S, Clape C, Iankova I *et al.* CXC ligand 5 is an adipose-tissue derived factor that links obesity to insulin resistance. Cell Metab. 9(4), 339-349 (2009).

101. Mei J, Liu Y, Dai N, Favara M, Greene T, Jeyaseelan S, Poncz M, Lee JS, Worthen GS. CXCL5 regulates chemokine scavenging and pulmonary host defense to bacterial infection. Immunity. 33(1), 106-117 (2010).

102. Pickens SR, Chamberlain ND, Volin MV, Gonzalez M, Pope RM, Mandelin AM, Kolls JK, Shahrara S. Anti-CXCL5 therapy ameliorates IL-17-induced arthritis by decreasing joint vascularization. Angiogenesis. 14(4), 443-455 (2011).

103. Koch AE, Kunkel SL, Harlow LA, Mazarakis DD, Haines GK, Burdick MD, Pope RM, Walz A, Strieter RM. Epithelial neutrophil activating peptide-78: a novel chemotactic cytokine for neutrophils in arthritis. J. Clin. Invest. 94(3), 1012-1018 (1994).

104. Dawes JM, Calvo M, Perkins JR, Paterson KJ, Kiesewetter H, Hobbs C, Kaan TK, Orengo C, Bennett DL, McMahon SB. CXCL5 mediates UVB irradiation-induced pain. Sci. Transl. Med. 3(90), 90ra60- (2011).

105. Li A, King J, Moro A, Sugi MD, Dawson DW *et al.* Overexpression of CXCL5 is associated with poor survival in patients with pancreatic cancer. Am. J. Pathol. 178(3), 1340-1349 (2011).

106. Linge HM, Collin M, Nordenfelt P, Morgelin M, Malmsten M, Egesten A. The human CXC chemokine granulocyte chemotactic protein 2 (GCP-2)/CXCL6 possesses membrane-disrupting properties and is antibacterial. Antimicrob. Agents Chemother. 52(7), 2599-2607 (2008).

107. Arenberg DA, Polverini PJ, Kunkel SL, Shanafelt A, Hesselgesser J, Horuk R, Strieter RM. The role of CXC chemokines in the regulation of angiogenesis in non-small cell lung cancer. J. Leukoc. Biol. 62(5), 554-562 (1997).

108. Bieche I, Chavey C, Andrieu C, Busson M, Vacher S, Le CL, Guinebretiere JM, Burlinchon S, Lidereau R, Lazennec G. CXC chemokines located in the 4q21 region are up-regulated in breast cancer. Endocr. Relat Cancer. 14(4), 1039-1052 (2007).

109. Xu L, Duda DG, di TE, Ancukiewicz M, Chung DC *et al.* Direct evidence that bevacizumab, an anti-VEGF antibody, up-regulates SDF1alpha, CXCR4, CXCL6, and neuropilin 1 in tumors from patients with rectal cancer. Cancer Res. 69(20), 7905-7910 (2009).

110. Keeley EC, Mehrad B, Strieter RM. Chemokines as mediators of tumor angiogenesis and neovascularization. Exp. Cell Res. 317(5), 685-690 (2011).

111. Harter L, Petersen F, Flad HD, Brandt E. Connective tissue-activating peptide III desensitizes chemokine receptors on neutrophils. Requirement for proteolytic formation of the neutrophil-activating peptide 2. J. Immunol. 153(12), 5698-5708 (1994).

112. Waugh DJ, Wilson C. The interleukin-8 pathway in cancer. Clin. Cancer Res. 14(21), 6735-6741 (2008).

113. Baldwin ET, Weber IT, St CR, Xuan JC, Appella E, Yamada M, Matsushima K, Edwards BF, Clore GM, Gronenborn AM, . Crystal structure of interleukin 8: symbiosis of NMR and crystallography. Proc. Natl. Acad. Sci. U. S. A. 88(2), 502-506 (1991).

114. Clore GM, Appella E, Yamada M, Matsushima K, Gronenborn AM. Three-dimensional structure of interleukin 8 in solution. Biochemistry. 29(7), 1689-1696 (1990).

115. Nesmelova IV, Sham Y, Dudek AZ, van Eijk LI, Wu G, Slungaard A, Mortari F, Griffioen AW, Mayo KH. Platelet factor 4 and interleukin-8 CXC chemokine heterodimer formation modulates function at the quaternary structural level. J. Biol. Chem. 280(6), 4948-4958 (2005).

116. Nesmelova IV, Sham Y, Gao J, Mayo KH. CXC and CC chemokines form mixed heterodimers: association free energies from molecular dynamics simulations and experimental correlations. J. Biol. Chem. 283(35), 24155-24166 (2008).

117. Brown AJ, Joseph PR, Sawant KV, Rajarathnam K. Chemokine CXCL7 Heterodimers: Structural Insights, CXCR2 Receptor Function, and Glycosaminoglycan Interactions. Int. J. Mol. Sci. 18(4) (2017).

118. Carlson J, Baxter SA, Dreau D, Nesmelova IV. The heterodimerization of platelet-derived chemokines. Biochim. Biophys. Acta. 1834(1), 158-168 (2013).

119. Fairbrother WJ, Reilly D, Colby TJ, Hesselgesser J, Horuk R. The solution structure of melanoma growth stimulating activity. J. Mol. Biol. 242(3), 252-270 (1994).

120. Kim KS, Clark-Lewis I, Sykes BD. Solution structure of GRO/melanoma growth stimulatory activity determined by 1H NMR spectroscopy. J. Biol. Chem. 269(52), 32909-32915 (1994).

121. Qian YQ, Johanson KO, McDevitt P. Nuclear magnetic resonance solution structure of truncated human GRObeta [5-73] and its structural comparison with CXC chemokine family members GROalpha and IL-8. J. Mol. Biol. 294(5), 1065-1072 (1999).

122. Shao W, Jerva LF, West J, Lolis E, Schweitzer BI. Solution structure of murine macrophage inflammatory protein-2. Biochemistry. 37(23), 8303-8313 (1998).

123. Rajasekaran D, Keeler C, Syed MA, Jones MC, Harrison JK, Wu D, Bhandari V, Hodsdon ME, Lolis EJ. A model of GAG/MIP-2/CXCR2 interfaces and its functional effects. Biochemistry. 51(28), 5642-5654 (2012).

124. Sepuru KM, Poluri KM, Rajarathnam K. Solution structure of CXCL5--a novel chemokine and adipokine implicated in inflammation and obesity. PLoS. One. 9(4), e93228- (2014).

125. Young H, Roongta V, Daly TJ, Mayo KH. NMR structure and dynamics of monomeric neutrophil-activating peptide 2. Biochem. J. 338 (Pt 3), 591-598 (1999).

126. Malkowski MG, Lazar JB, Johnson PH, Edwards BF. The amino-terminal residues in the crystal structure of connective tissue activating peptide-III (des10) block the ELR chemotactic sequence. J. Mol. Biol. 266(2), 367-380 (1997).

127. Eigenbrot C, Lowman HB, Chee L, Artis DR. Structural change and receptor binding in a chemokine mutant with a rearranged disulfide: X-ray structure of E38C/C50AIL-8 at 2 A resolution. Proteins. 27(4), 556-566 (1997).

128. Gerber N, Lowman H, Artis DR, Eigenbrot C. Receptor-binding conformation of the "ELR" motif of IL-8: X-ray structure of the L5C/H33C variant at 2.35 A resolution. Proteins. 38(4), 361-367 (2000).

129. Skelton NJ, Quan C, Reilly D, Lowman H. Structure of a CXC chemokine-receptor fragment in complex with interleukin-8. Structure. 7(2), 157-168 (1999).

130. Baldwin ET, Weber IT, St CR, Xuan JC, Appella E, Yamada M, Matsushima K, Edwards BF, Clore GM, Gronenborn AM, . Crystal structure of interleukin 8: symbiosis of NMR and crystallography. Proc. Natl. Acad. Sci. U. S. A. 88(2), 502-506 (1991).

131. Birgit Goger, Yvonne Halden, Angelika Rek, Roland Mösl, David Pye, John Gallagher, and Andreas J.Kung. Different Affinities of Glycosaminoglycan Oligosaccharides for Monomeric and Dimeric Interleukin-8:? A Model for Chemokine Regulation at Inflammatory Sites. Biochemistry41 (5), 1640-1646 (2002)

132. Gandhi NS, Mancera RL. Molecular dynamics simulations of CXCL-8 and its interactions with a receptor peptide, heparin fragments, and sulfated linked cyclitols. J. Chem. Inf. Model. 51(2), 335-358 (2011).

133. Spillmann D, Witt D, Lindahl U. Defining the interleukin-8-binding domain of heparan sulfate. J. Biol. Chem. 273(25), 15487-15493 (1998).

134. Bitomsky W. and Wade R.C. Docking of glycosaminoglycans to heparin-binding proteins. Validation for aFGF, bFGF, and antithrombin and applications to IL-8. J. Am. Chem. Soc.121, 3004-3013 (1999)

135. Lortat-Jacob H, Grosdidier A, Imberty A. Structural diversity of heparan sulfate binding domains in chemokines. Proc. Natl. Acad. Sci. U. S. A. 99(3), 1229-1234 (2002).

136. Poluri KM, Joseph PR, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. J. Biol. Chem. 288(35), 25143-25153 (2013).

137. Brown AJ, Sepuru KM, Rajarathnam K. Structural Basis of Native CXCL7 Monomer Binding to CXCR2 Receptor N-Domain and Glycosaminoglycan Heparin. Int. J. Mol. Sci. 18(3) (2017).

138. Sepuru KM, Nagarajan B, Desai UR, Rajarathnam K. Molecular Basis of Chemokine CXCL5-Glycosaminoglycan Interactions. J. Biol. Chem. 291(39), 20539-20550 (2016).

139. Pei J. Multiple protein sequence alignment. Curr. Opin. Struct. Biol. 18(3), 382-386 (2008).

140. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol Evol. 30(12), 2725-2729 (2013).

141. Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR. SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. Syst. Biol. 61(1), 90-106 (2012).

142. Hohna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. Syst. Biol. 65(4), 726-736 (2016).

143. Didelot X, Fraser C, Gardy J, Colijn C. Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks. Mol. Biol Evol. 34(4), 997-1007 (2017).

144. Li T, Lai XL, Zhong Y. [Introduction to the methods of constructing phylogenetic trees with DNA sequences]. Yi. Chuan. 26(2), 205-210 (2004).

145. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol Evol. 17(1), 32-43 (2000).

146. Smith GP, Hood L, Fitch WM. Antibody diversity. Annu. Rev. Biochem. 40, 969-1012 (1971).

147. Smith GP. Unequal crossover and the evolution of multigene families. Cold Spring Harb. Symp. Quant. Biol. 38, 507-513 (1974).

148. Horner DS, Pirovano W, Pesole G. Correlated substitution analysis and the prediction of amino acid structural contacts. Brief. Bioinform. 9(1), 46-56 (2008).

149. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino BC. MISTIC: Mutual information server to infer coevolution. Nucleic Acids Res. 41(Web Server issue), W8-14 (2013).

150. Floudas CA. Computational methods in protein structure prediction. Biotechnol Bioeng. 97(2), 207-213 (2007).

151. Xiang Z. Advances in homology protein structure modeling. Curr. Protein Pept. Sci. 7(3), 217-227 (2006).

152. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 22(2), 195-201 (2006).

153. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci. Chapter 2, Unit- (2007).

154. Eyrich VA, Rost B. META-PP: single interface to crucial prediction servers. Nucleic Acids Res. 31(13), 3308-3310 (2003).

155. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32(Web Server issue), W526-W531 (2004).

156. LATHE GH, RUTHVEN CR. The separation of substances and estimation of their relative molecular sizes by the use of colums of starch in water. Biochem. J. 62(4), 665-674 (1956).

157. Barth HG, Boyes BE, Jackson C. Size exclusion chromatography. Anal. Chem. 66(12), 595R-620R (1994).

158. Williams A, Frasca V. Ion-exchange chromatography. Curr. Protoc. Protein Sci. Chapter 8, Unit8- (2001).

159. Proudfoot AE, Fritchley S, Borlat F, Shaw JP, Vilbois F, Zwahlen C, Trkola A, Marchant D, Clapham PR, Wells TN. The BBXB motif of RANTES is the principal site for heparin binding and controls receptor selectivity. J Biol Chem. 276(14), 10620-10626 (2001).

160. Lau EK, Paavola CD, Johnson Z, Gaudry JP, Geretti E, Borlat F, Kungl AJ, Proudfoot AE, Handel TM. Identification of the glycosaminoglycan binding site of the CC chemokine, MCP-1: implications for structure and function in vivo. J Biol Chem. 279(21), 22294-22305 (2004).

161. Beychok S. Circular dichroism of biological macromolecules. Science. 154(3754), 1288-1299 (1966).

162. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1(6), 2876-2890 (2006).

163. Weber G. Fluorescence in biophysics: accomplishments and deficiencies. Methods Enzymol. 278, 1-15 (1997).

164. Clore GM, Schwieters CD. Theoretical and computational advances in biomolecular NMR spectroscopy. Curr. Opin. Struct. Biol. 12(2), 146-153 (2002).

165. Siegal G, van DJ, Baldus M. Biomolecular NMR: recent advances in liquids, solids and screening. Curr. Opin. Chem. Biol. 3(5), 530-536 (1999).

166. Goto NK, Kay LE. New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr. Opin. Struct. Biol. 10(5), 585-592 (2000).

167. Kainosho M. Isotope labelling of macromolecules for structural determinations. Nat. Struct. Biol. 4 Suppl, 858-861 (1997).

168. Srivastava S, Roy R, Gupta V, Tiwari S, Srivastava AN, and Sonkar AA. Proton HR-MAS MR Spectroscopy of Oral Squamous Cell Carcinoma Tissues: An ex-vivo study to Identify Malignancy Induced Metabolic Fingerprints. Metabolomics7, 278- (2011)

169. Srivastava S, Sonkar AA, and Roy R. Oral Squamous Cell Carcinoma: Insights with metabonomics. Chemistry & Biology Interface2, 206-

170. Sidhu OP, Annarao S, Chatterjee S, Tuli R, Roy R, and Khetrapal CL. Metabolic alterations of Withania somnifera (L.) Dunal fruits at different development stages by NMR Spectroscopy. Phytochemcal Analysis22 (6), 492-502 (2011)

171. Bharti SK, Jaiswal V, Ghoshal UC, Baijal SS, Ghoshal C, Roy R, and Khetrapal CL. Metabolomic profiling of amoebic and pyogenic liver abscesses: An in-vitro NMR study. Metabolomics8 (4), 540-555 (2011)

172. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J. Magn Reson. 160(1), 65-73 (2003).

173. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. 273(1), 283-298 (1997).

174. Guntert P. Automated NMR structure calculation with CYANA. Methods Mol. Biol. 278, 353-378 (2004).

175. Kay LE, Gardner KH. Solution NMR spectroscopy beyond 25 kDa. Curr. Opin. Struct. Biol. 7(5), 722-731 (1997).

176. Otomo T, Teruya K, Uegaki K, Yamazaki T, Kyogoku Y. Improved segmental isotope labeling of proteins and application to a larger protein. J. Biomol. NMR. 14(2), 105-114 (1999).

177. Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. Proc. Natl. Acad. Sci. U. S. A. 95(12), 6705-6710 (1998).

178. Xu R, Ayers B, Cowburn D, Muir TW. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. Proc. Natl. Acad. Sci. U. S. A. 96(2), 388-393 (1999).

179. Liu D, Xu R, Cowburn D. Segmental isotopic labeling of proteins for nuclear magnetic resonance. Methods Enzymol. 462, 151-175 (2009).

180. Skrisovska L, Schubert M, Allain FH. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. J Biomol. NMR. 46(1), 51-65 (2010).

181. Tong KI, Yamamoto M, Tanaka T. A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. J Biomol. NMR. 42(1), 59-67 (2008).

182. Jaipuria G, Krishnarjuna B, Mondal S, Dubey A, Atreya HS. Amino acid selective labeling and unlabeling for protein resonance assignments. Adv. Exp. Med. Biol. 992, 95-118 (2012).

183. Kelly MJ, Krieger C, Ball LJ, Yu Y, Richter G, Schmieder P, Bacher A, Oschkinat H. Application of amino acid type-specific 1H- and 14N-labeling in a 2H-, 15N-labeled background to a 47 kDa homodimer: potential for NMR structure determination of large proteins. J Biomol. NMR. 14(1), 79-83 (1999).

184. Krishnarjuna B, Jaipuria G, Thakur A, D'Silva P, Atreya HS. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. J Biomol. NMR. 49(1), 39-51 (2011).

185. Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE. Selective methyl group protonation of perdeuterated proteins. J Mol. Biol. 263(5), 627-636 (1996).

186. Jaipuria G, Lobo NP, Shet D, Atreya HS. High resolution methyl selective (1)(3)C-NMR of proteins in solution and solid state. J Biomol. NMR. 54(1), 33-42 (2012).

187. Atreya HS, Chary KV. Selective 'unlabeling' of amino acids in fractionally 13C labeled proteins: an approach for stereospecific NMR assignments of CH3 groups in Val and Leu residues. J Biomol. NMR. 19(3), 267-272 (2001).

188. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei OA, Guntert P. Optimal isotope labelling for NMR protein structure determinations. Nature. 440(7080), 52-57 (2006).

189. Hanudatta S.Atreya. NMRmethods for fast data acquisition. Journal of the Indian Institute of Science90, 1- (2010)

190. Carlomagno T. Ligand-target interactions: what can we learn from NMR? Annu. Rev. Biophys. Biomol. Struct. 34, 245-266 (2005).

191. Pellecchia M. Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. Chem. Biol. 12(9), 961-971 (2005).

192. Zuiderweg ER. Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry. 41(1), 1-7 (2002).

193. Shimada I. NMR techniques for identifying the interface of a larger protein-protein complex: cross-saturation and transferred cross-saturation experiments. Methods Enzymol. 394, 483-506 (2005).

194. Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada I. A novel NMR method for determining the interfaces of large protein-protein complexes. Nat. Struct. Biol. 7(3), 220-223 (2000).

195. Bagga P, Chugani AN, Varadarajan KS, Patel AB. In vivo NMR studies of regional cerebral energetics in MPTP model of Parkinson's disease: recovery of cerebral metabolism with acute levodopa treatment. J Neurochem. 127(3), 365-377 (2013).

196. Bhattacharya S, Sheikh L, Tiwari V, Ghosh M, Patel JN, Patel AB, Nayar S. Protein-polymer functionalized aqueous ferrofluids showing high T2 relaxivity. J Biomed. Nanotechnol. 10(5), 811-819 (2014).

197. Chugh J, Chatterjee A, Kumar A, Mishra RK, Mittal R, Hosur RV. Structural characterization of the large soluble oligomers of the GTPase effector domain of dynamin. FEBS J. 273(2), 388-397 (2006).

198. Kashyap M and Bhavesh NS. Structural characterization of the RNA binding domain of human stem loop binding protein. J. Prot. Proteom.42, 109-114 (2017)

199. Kashyap M, Ganguly AK, Bhavesh NS. Structural delineation of stem-loop RNA binding by human TAF15 protein. Sci. Rep. 5, 17298- (2015).

200. Mandal PK, Bhavesh NS, Chauhan VS, Fodale V. NMR investigations of amyloid-beta peptide interactions with propofol at clinically relevant concentrations with and without aqueous halothane solution. J Alzheimers. Dis. 21(4), 1303-1309 (2010).

201. Munshi SU, Rewari BB, Bhavesh NS, Jameel S. Nuclear magnetic resonance based profiling of biofluids reveals metabolic dysregulation in HIV-infected persons and those on anti-retroviral therapy. PLoS. One. 8(5), e64298- (2013).

202. Trivedi DK, Bhatt H, Pal RK, Tuteja R, Garg B, Johri AK, Bhavesh NS, Tuteja N. Structure of RNA-interacting cyclophilin A-like protein from Piriformospora indica that provides salinity-stress tolerance in plants. Sci. Rep. 3, 3001- (2013).

203. Zhang Z, Keramisanou D, Dudhat A, Pare M, Gelis I. The C-terminal domain of human Cdc37 studied by solution NMR. J Biomol. NMR. 63(3), 315-321 (2015).

204. Malik N, Kumar A. Resonance assignment of disordered protein with repetitive and overlapping sequence using combinatorial approach reveals initial structural propensities and local restrictions in the denatured state. J Biomol. NMR. 66(1), 21-35 (2016).

205. Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. Curr. Opin. Struct. Biol. 32, 113-122 (2015).

206. Goto NK, Kay LE. New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr. Opin. Struct. Biol. 10(5), 585-592 (2000).

207. Sattler M, Fesik SW. Use of deuterium labeling in NMR: overcoming a sizeable problem. Structure. 4(11), 1245-1249 (1996).

208. Keramisanou D, Aboalroub A, Zhang Z, Liu W, Marshall D, Diviney A, Larsen RW, Landgraf R, Gelis I. Molecular Mechanism of Protein Kinase Recognition and Sorting by the Hsp90 Kinome-Specific Cochaperone Cdc37. Mol. Cell. 62(2), 260-271 (2016).

209. Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell. 131(4), 756-769 (2007).

210. Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn Reson. Spectrosc. 73, 1-16 (2013).

211. Otting G, Wuthrich K. Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. Q. Rev. Biophys. 23(1), 39-96 (1990).

212. Ferrage F, Dutta K, Shekhtman A, Cowburn D. Structural determination of biomolecular interfaces by nuclear magnetic resonance of proteins with reduced proton density. J Biomol. NMR. 47(1), 41-54 (2010).

213. Hall DA, Vander Kooi CW, Stasik CN, Stevens SY, Zuiderweg ER, Matthews RG. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. Proc. Natl. Acad. Sci. U. S. A. 98(17), 9521-9526 (2001).

214. Nakanishi T, Miyazawa M, Sakakura M, Terasawa H, Takahashi H, Shimada I. Determination of the interface of a large protein complex by transferred cross-saturation measurements. J Mol. Biol. 318(2), 245-249 (2002).

215. Baxter NJ, Williamson MP. Temperature dependence of 1H chemical shifts in proteins. J. Biomol. NMR. 9(4), 359-369 (1997).

216. Baxter NJ, Hosszu LL, Waltho JP, Williamson MP. Characterisation of low free-energy excited states of folded proteins. J Mol. Biol. 284(5), 1625-1639 (1998).

217. Tunnicliffe RB, Waby JL, Williams RJ, Williamson MP. An experimental investigation of conformational fluctuations in proteins G and L. Structure. 13(11), 1677-1684 (2005).

218. Williamson MP. Many residues in cytochrome c populate alternative states under equilibrium conditions. Proteins. 53(3), 731-739 (2003).

219. Kumar A, Srivastava S, Hosur RV. NMR characterization of the energy landscape of SUMO-1 in the native-state ensemble. J Mol. Biol. 367(5), 1480-1493 (2007).

220. Englander SW, Krishna MM. Hydrogen exchange. Nat. Struct. Biol. 8(9), 741-742 (2001).

221. Hoang L, Bedard S, Krishna MM, Lin Y, Englander SW. Cytochrome c folding pathway: kinetic native-state hydrogen exchange. Proc. Natl. Acad. Sci. U. S. A. 99(19), 12173-12178 (2002).

222. Thakur A, Chandra K, Dubey A, D'Silva P, Atreya HS. Rapid characterization of hydrogen exchange in proteins. Angew. Chem. Int. Ed Engl. 52(9), 2440-2443 (2013).

223. Krishna MM, Hoang L, Lin Y, Englander SW. Hydrogen exchange methods to study protein folding. Methods. 34(1), 51-64 (2004).

224. Bruschweiler R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. Curr. Opin. Struct. Biol. 13(2), 175-183 (2003).

225. Kumar A, Srivastava S, Kumar MR, Mittal R, Hosur RV. Residue-level NMR view of the urea-driven equilibrium folding transition of SUMO-1 (1-97): native preferences do not increase monotonously. J Mol. Biol. 361(1), 180-194 (2006).

226. Kay LE. Protein dynamics from NMR. Nat. Struct. Biol. 5 Suppl, 513-517 (1998).

227. Kumar A, Srivastava S, Mishra RK, Mittal R, Hosur RV. Local structural preferences and dynamics restrictions in the urea-denatured state of SUMO-1: NMR characterization. Biophys. J. 90(7), 2498-2509 (2006).

228. Ishima R, Torchia DA. Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. J Biomol. NMR. 25(3), 243-248 (2003).

229. Takahashi D, Hiromasa Y, Kim Y, Anbanandam A, Yao X, Chang KO, Prakash O. Structural and dynamics characterization of norovirus protease. Protein Sci. 22(3), 347-357 (2013).

230. Cook GA, Pajewski R, Aburi M, Smith PE, Prakash O, Tomich JM, Gokel GW. NMR structure and dynamic studies of an anion-binding, channel-forming heptapeptide. J Am. Chem. Soc. 128(5), 1633-1638 (2006).

231. C.S.Johnson Jr. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. Progress in Nuclear Magnetic Resonance Spectroscopy34, 203-256 (1999)

232. Luan J, Shattuck-Brandt R, Haghnegahdar H, Owen JD, Strieter R, Burdick M, Nirodi C, Beauchamp D, Johnson KN, Richmond A. Mechanism and biological significance of constitutive expression of MGSA/GRO chemokines in malignant melanoma tumor progression. J Leukoc. Biol. 62(5), 588-597 (1997).

233. Anisowicz A, Bardwell L, Sager R. Constitutive overexpression of a growth-regulated gene in transformed Chinese hamster and human cells. Proc. Natl. Acad. Sci. U. S. A. 84(20), 7188-7192 (1987).

234. Richmond A, Balentien E, Thomas HG, Flaggs G, Barton DE, Spiess J, Bordoni R, Francke U, Derynck R. Molecular characterization and chromosomal mapping of

melanoma growth stimulatory activity, a growth factor structurally related to beta-thromboglobulin. EMBO J. 7(7), 2025-2033 (1988).

235. Haskill S, Peace A, Morris J, Sporn SA, Anisowicz A, Lee SW, Smith T, Martin G, Ralph P, Sager R. Identification of three related human GRO genes encoding cytokine functions. Proc. Natl. Acad. Sci. U. S. A. 87(19), 7732-7736 (1990).

236. Smith DF, Galkina E, Ley K, Huo Y. GRO family chemokines are specialized for monocyte arrest from flow. Am. J Physiol Heart Circ. Physiol. 289(5), H1976-H1984 (2005).

237. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, Rousseau S, Eidelman DH, Hamid Q. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. J Immunol. 191(5), 2731-2741 (2013).

238. Doll D, Keller L, Maak M, Boulesteix AL, Siewert JR, Holzmann B, Janssen KP. Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. Int. J Colorectal Dis. 25(5), 573-581 (2010).

239. Kocher AA, Schuster MD, Bonaros N, Lietz K, Xiang G *et al.* Myocardial homing and neovascularization by human bone marrow angioblasts is regulated by IL-8/Gro CXC chemokines. J Mol. Cell Cardiol. 40(4), 455-464 (2006).

240. Luttichau HR. The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist. J Biol. Chem. 285(12), 9137-9146 (2010).

241. Sajadi SM, Khoramdelazad H, Hassanshahi G, Rafatpanah H, Hosseini J, Mahmoodi M, Arababadi MK, Derakhshan R, Hasheminasabzavareh R, Hosseini-Zijoud SM, Ahmadi Z. Plasma levels of CXCL1 (GRO-alpha) and CXCL10 (IP-10) are elevated in type 2 diabetic patients: evidence for the involvement of inflammation and angiogenesis/angiostasis in this disease state. Clin. Lab. 59(1-2), 133-137 (2013).

242. Fimmel S, Devermann L, Herrmann A, Zouboulis C. GRO-alpha: a potential marker for cancer and aging silenced by RNA interference. Ann. N. Y. Acad. Sci. 1119, 176-189 (2007).

243. Felice TIRONE, Sebastiano Cavallaro, Stefano FARIOLI-VECCHIONI, Laura MICHELI, Luca Leonardi, Irene CINA, and Manuela CECCARELLI. Cxcl3 chemokine for the therapeutic treatment of medulloblastoma.EP 2903633 A1 (2015)

Chapter 2: Mechanistic Insights into Molecular Evolution of Species Specific Differential Glycosaminoglycan Binding Surfaces in GRO Chemokines

Abstract

Unraveling the evolutionary concepts of GRO chemokines is essential to understand its impact on their differential structural and functional features. In the current chapter, comprehensive evolutionary analysis of GRO genes among different mammalian species was carried out using a combination of various bioinformatic tools including phylogenetic, selection, substitution rate, and electrostatic surface analysis. Phylogenetic analysis illustrated a species specific evolution pattern. Selection analysis evidenced that these genes have undergone concerted evolution. Seventeen positively selected sites were obtained, although majority of the proteins are under purifying selection. Interestingly, these positively selected sites are more concentrated on the C-terminal/GAG binding and dimerization segment compared to their receptor binding domain. Substitution rate analysis confirmed the C-terminal domain of GRO genes as the highest substituted segment. Further, structural analysis established that the nucleotide alternations in the GAG binding domain are the source of surface charge modulation, thus generating the differential and multiple GAG binding surfaces as per evolutionary pressure, although the helical surface is primordial for GAG binding. Indeed, such variable electrostatic surfaces are crucial to regulate chemokine gradient formation during host's defense against pathogens and also explain the significance of chemokine promiscuity.

2.1. Introduction

GRO chemokines (GRO α, Gro β, and GRO γ) belong to the NAC subfamily of CXC chemokines [1]. NACs share a structurally conserved monomeric fold, comprising of long

disordered N-terminal domain containing signature ELR (Glu-Leu-Arg) motif followed by a 3_{10} helix, three anti parallel β-strands and a C-terminal α-helix **(Fig. 2.1 A-B)**. NACs form CXC-type dimer which is formed by the antiparallel arrangement of two C-terminal α helices on the top of six stranded antiparallel β-sheet **(Fig. 2.1 C)**. The major hallmarks for CXC dimerization involve the formation of six backbone hydrogen bonds between the β_1–β'_1strands of both the monomeric units **(Fig. 2.1 D)**. Several other intermolecular contacts for CXC dimer formation includes hydrophobic, electrostatic/Van der Waals interactions between the α-helices (α-α') and with β-strand (α-β') residues as shown in **Fig. 2.1 D-E**. N-terminal residues including E6-R8 and L12-K21 constitutes the receptor binding domain of NACs and conserved GAG binding residues (H19, K21, K45, K61, K65) are distributed on the helices (α-domain) **(Fig. 2. 1 F)** [2]. Further, a unique GAG binding β-domain surface comprising of residues R8, R48, K49, K29 in human CXCL1 chemokine is reported **(Fig. 2.1 F)** [3].

Although, many reports have exposed various evolutionary perspectives about different NACs, still there is a lag in tracing their structural and functional implications using evolutionary concepts. To throw light on the evolutionary characteristics of NACs, the evolution of GRO chemokines was studied. GRO chemokines are highly related to each other both in terms of sequence and structure, but they are involved in different functions like differential receptor activation, binding, and expression patterns [4-6]. This can be attributed to their differential homo/hetero oligomerization, binding to receptors/GAGs, or a combination of such molecular recognition events [4].

In this chapter, the evolutionary mechanisms that tune the structural and functional aspects of GRO chemokines were delineated. GRO genes across different mammalian species were analyzed using phylogenetic, selection, substitution rate analysis, conservation scores, nucleotide alterations, and electrostatic surface potentials.

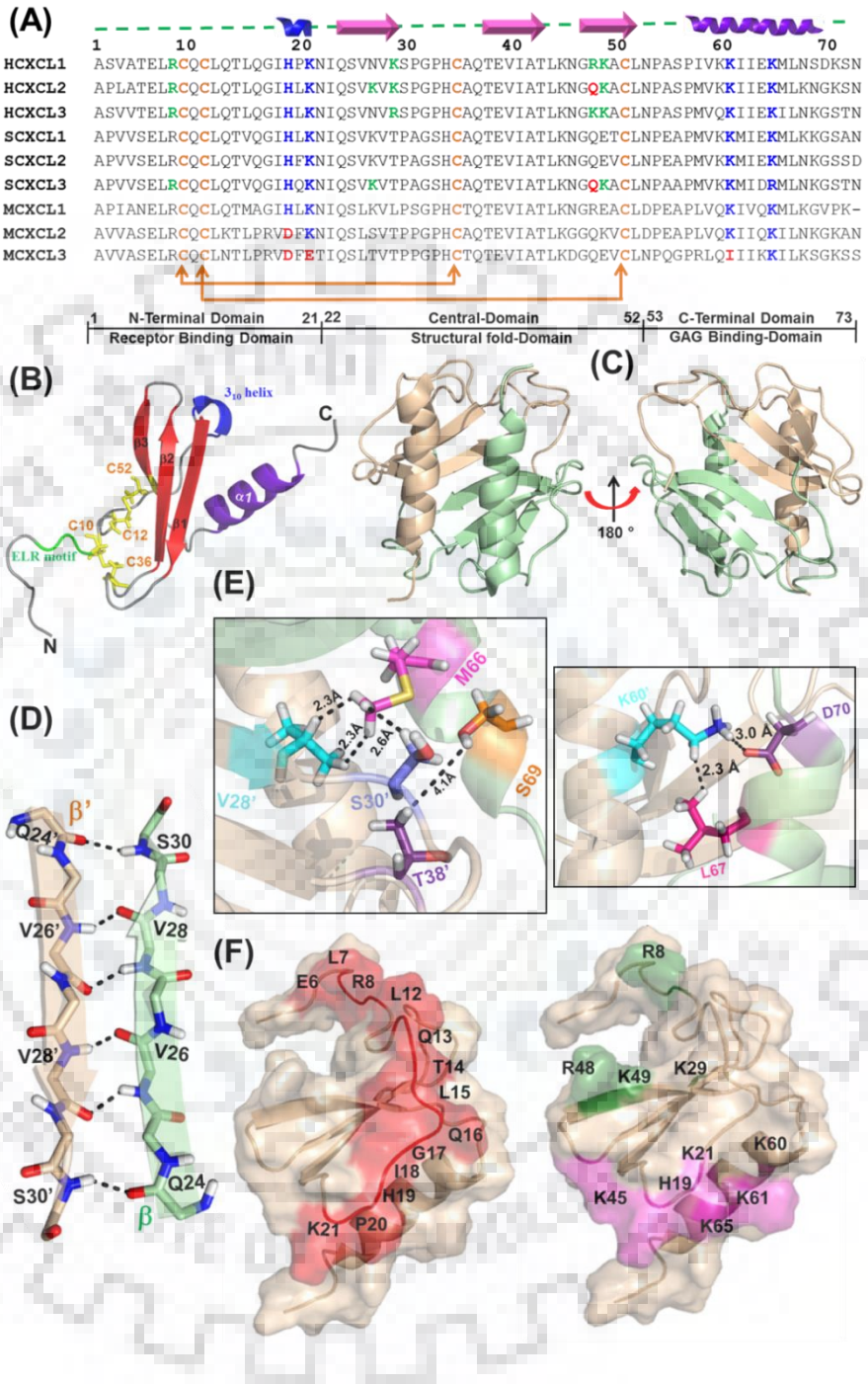


Figure 2.1: (A) *Amino acid sequences of GRO family of chemokines from human (H), horse (S), and murine (M). Residues involved in canonical α-helical surface GAG binding are marked in blue and specific to β-sheet surface GAG binding are marked in green, charge alteration among these residues is highlighted with red.* **(B)** *Structural elements in the monomeric structure of ELR-CXC chemokines.* **(C)** *Dimeric structure of HCXCL1.* **(D)** *Backbone H-bonding of β_1-β'_1 residues*

dimer interface. **(E)** *Essential dimer interface contacts between C-terminal helices and with β-strands.* **(F)** *Surface structure of HCXCL1 monomer marked with receptor binding residues (red) and GAG binding residues forming α-domain (pink) and β-domain (green).*

2.2 Materials and methods

2.2.1 Sequence alignment and phylogenetic analysis

A total of 83 gene/amino acid sequences of GRO chemokines (CXCL1-38, CXCL2- 21, CXCL3- 24) from different mammalian species were obtained from the sequence databases of GENBANK (http://www.ncbi.nlm.nih.gov/genbank/), Uniprot (http://www.uniprot.org/) and Ensemble (http://www.ensembl.org/), for which the unique IDs are listed in **Appendix I**.

Out of these sequences, all three GRO gene sequences were found for the following species: primates (human, chimpanzee, orangutan, gibbon, crab eating macaque), lagomorpha (rabbit), rodents (rat, murine, chinese hamster), and laurasiatheria (horse, cow). Only two gene sequences for walrus, bison, white rhinoceros, pig, and buffalo of laurasiatheria were used. Although evolutionary analysis was performed by using the complete set of 83 sequences, the species containing either 2 or 3 GRO genes were used for all the analysis related to gene duplication phenomenon.

Multiple sequence alignment for the nucleotide sequences of GRO chemokines was generated using CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) by employing default settings for gap extension and gap opening penalty [7]. Multiple sequence alignment thus obtained was used as an input data for molecular evolutionary genetic analysis (MEGA 6) software [8]. Phylogenetic tree of GRO family chemokines among different mammalian species constructed using neighbor joining (NJ) method based on p-distance in MEGA 6 (molecular evolutionary genetic analysis) [8]. Reliability of the tree was assessed by bootstrap method using 1000 bootstrap replications that resulted in bootstrap proportion for each internal branch in the tree.

2.2.2 Selection analysis

Multiple sequence alignment of GRO nucleotide sequences was screened for the gene recombination using genetic algorithm recombination detection (GARD) approach of

DATAMONKEY server [9,10]. In addition, GENECONV method of the RDP4 program was used to identify the gene conversion events [11].

Codon based maximum likelihood method was employed to test positive selection in GRO genes and to infer amino acid sites under the positive selection during evolution using DATAMONKEY server and codeml program in the PAML 4.9a [10,12]. DATAMONKEY programs include: Single likelihood ancestor counting (SLAC), fixed effect likelihood (FEL), internal fixed effects likelihood (IFEL), mixed effect model evolution (MEME), branch site random effect likelihood (REL), fast unbiased bayesian approximation (FUBAR) [10,13-15]. For SLAC, FEL, iFEL and MEME methods, the P-value of 0.1 was used. P-value was set to 0.95 for FUBAR, and the Bayes factor of > 95 was used for REL. For codeml program, once the positive selection was confirmed, Naive or Bayes Empirical Bayes (NEB/BEB) approach was used to calculate the posterior probability for all the sites evolved under positive selection [16]. As reported previously, the resulting positively selected amino acid sites that were obtained at least by two methods were considered [17-21]. Further, to examine the nature of duplication (concerted/divergent) among GRO genes, the pair wise values of dN and dS were compared for all the species.

2.2.3 Conservation score

To determine the site specific conservation score in the GRO proteins, ConSurf server was used. Conservation score or the percentage of occurrence of specific amino acid at particular site was calculated using empirical Bayesian method based on multiple sequence alignment [22]. Conservation profile for each of the site was also generated using WebLogo [23].

2.2.4 Substitution rates

Nucleotide substitution rates for GRO genes were determined using p-distance model (assumes number of nucleotide changes per site) in MEGA 6.0. p-distance is the proportion (*p*) of nucleotide sites at which two sequences being compared are distinct. It is calculated by dividing the number of nucleotide differences by the total number of nucleotides being compared [8].

2.2.5 Computation of pair wise omega values among different species

To assess the variations among the non-synonymous/synonymous substitution rate ratios among different GRO domains, pair wise omega values were computed for three GRO domain sequences independently using Nei–Gojobori (NG) [24] method with JCODA [25].

2.2.6 Coevolution analysis

Coevolving residues in the GRO genes were predicted using MISTIC (mutual information server to infer coevolution) server. It is based on mutual coevolutionary relationship between the amino acids that play an essential functional and structural role in the protein family [26]. Analysis was done using the available structures of HCXCL1, HCXCL2 and MCXCL2. Average mutual information (AMI) score was calculated for all the coevolving pairs of positively selected residues.

2.2.7 Sequence and structural analysis of GRO chemokines

The nucleotide codon vs amino acid alterations of the CXCL1-3 sequences for all the species present in the phylogenetic tree were analyzed independently by aligning amino acids with nucleotide sequences. Structural models for human CXCL3, murine CXCL3, and horse CXCL1/CXCL2/CXCL3 were generated through homology modeling based on target template alignment using Promod II in Swiss model server by employing known GRO structures as templates **(Appendix II)** [27]. Homodimers were generated through the monomeric counterparts of CXCL1/CXCL2/CXCL3 structures by performing symmetry operations employing PyMOL macros using Cα chains and dimer interface residues as reference restraints. Electrostatic surface potential maps for human, horse, and murine CXCL1-3 monomeric and dimeric chemokines were generated using vaccum electrostatics of PyMOL molecular graphic system [28].

2.3 Results

2.3.1. GRO genes adapted species specific evolutionary patterns

Phylogenetic analysis exploiting the sequence data enriches with an accurate picture of how different genes evolve, and are related to each other with respect to time and speciation. It not only provides the information about the gene evolution pattern at present, but also predicts the patterns

of evolution that can be followed in future. To rejuvenate the historical events that have occurred during the course of evolution of GRO genes, phylogenetic tree using the amino acid sequences of GRO proteins was constructed **(Fig. 2.2)**.

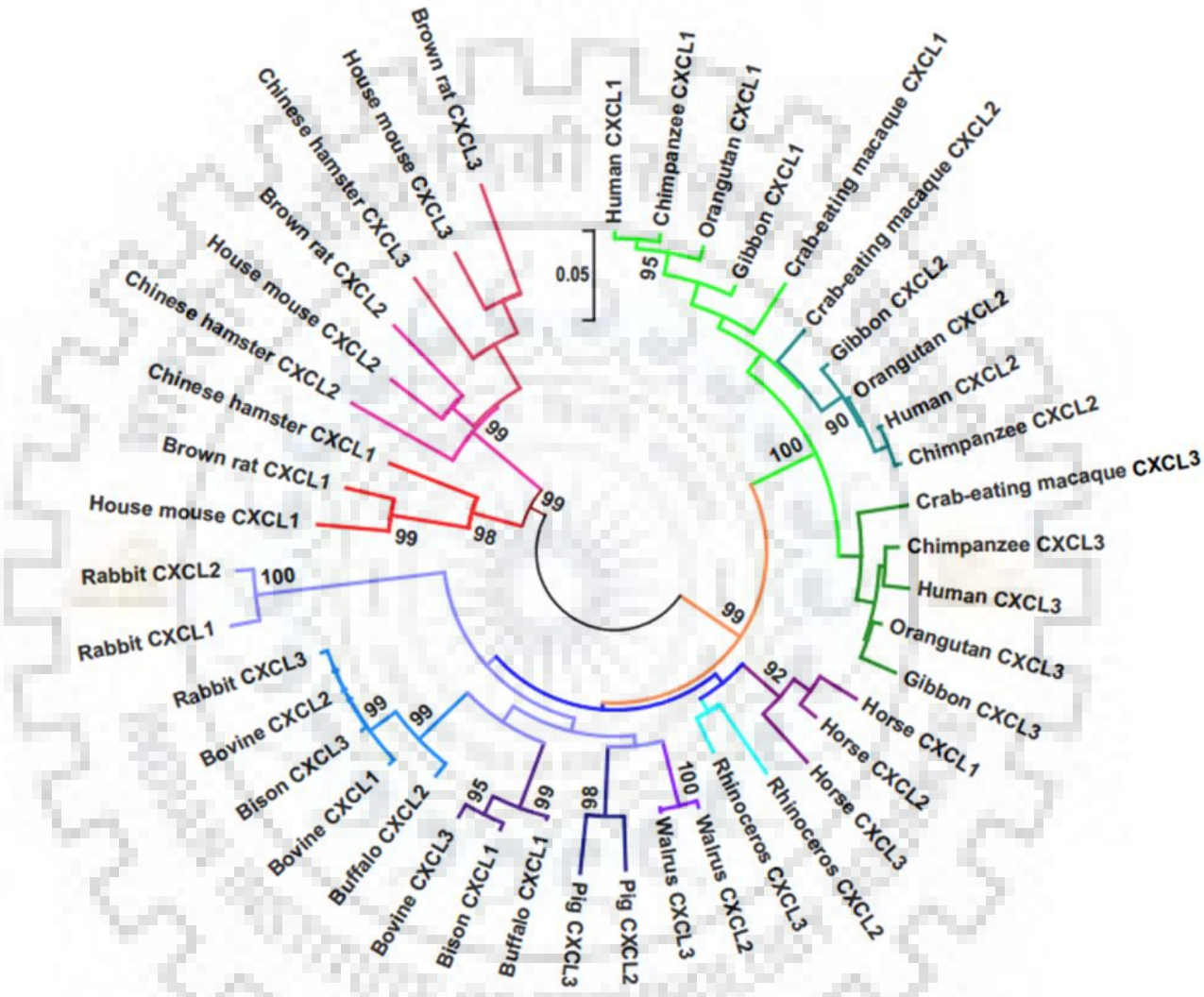


Figure 2.2: *Phylogenetic tree of GRO family chemokines among different mammalian species. Scale bar represents distance that corresponds to 5 % amino acid differences between the sequences*

Broadly, the tree topology displayed three major clades depending on the GRO genes belonging to different suites of species namely; primates, laurasiatheria, and rodents. Distribution of these nodes in the tree unraveled a distant relationship among GRO genes. Moving towards the

inner branches, the tree has unlocked the trends of their inherent duplication and speciation events. Within primates, CXCL1, CXCL2, and CXCL3 are more closely related to their orthologs. A similar topology was observed in rodents, implying duplication as a pre-speciation event. A contrasting trend was marked in case of laurasiatheria, in which paralogs of GRO chemokines are more closely related than to their orthologs.

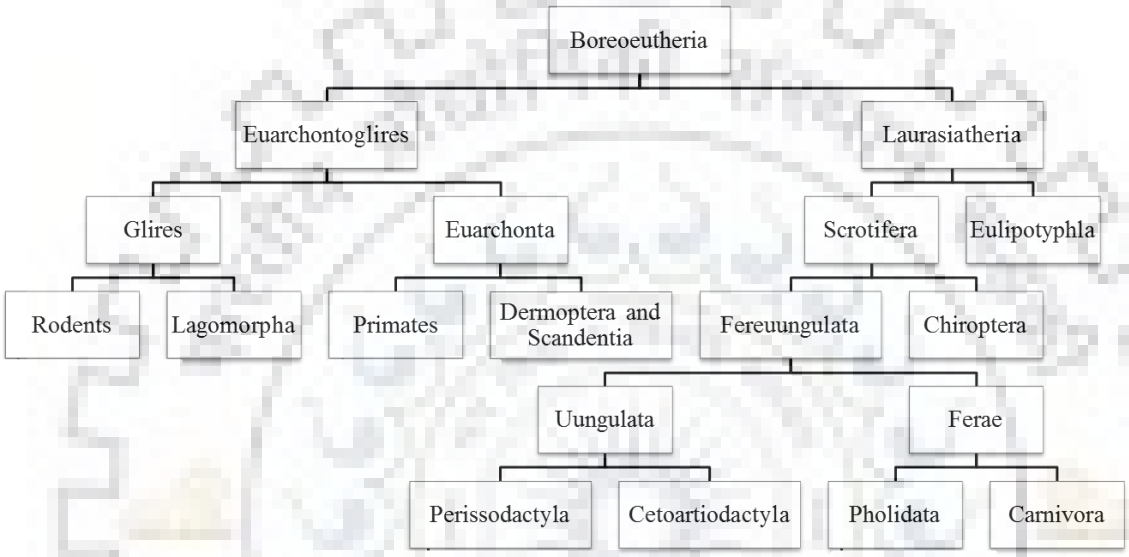


Figure 2.3: *Schematic showing a brief evolutionary summary of mammalian species*

Further, to establish the origin of such varied phylogenetic features of the GRO chemokines in mammalian species, their ancestral origin and classification was traced **(Fig. 2.3)**. Boroeutheria with two superorder branches euarchontoglires and laurasiatheria, were considered to analyze the distinguishable patterns of GRO chemokine evolution. The primates, lagomorpha, and rodents belong to the superorder euarchontoglires, where they show close relationship to their paralogs implicating that duplication of GRO genes in euarchontoglires is pre-speciation event and confirms its ancestral origin. In contrast to euarchontoglires, the superorder laurasiatheria comprising of perissodactyla, cetoartiodactyla, and carnivora species showed more similarity among the paralogs of GRO genes implying that the duplication of GRO genes in ferreungulata is probably a post-speciation event. Identification of the ancestral origin of laurasiatheria with current set of data of GRO chemokines is not feasible. Full genome annotation of more mammalian species under

laurasiatheria is required to reconsider their evolutionary patterns and to consider magnoorder boreoeutheria as the most probable ancestor for duplicated GRO gene family.

Moving towards the other panel of the tree, one alluring difference was marked in case of rabbit GRO genes. Despite of rabbit being a species of euarchontoglires, it forms a clade with laurasiatheria. Alarmingly, it has been observed that within rabbit GRO genes, CXCL3 shows a different behavior and is close to bovine CXCL2 and forms a single node. In contrast, the other two paralogs of rabbit (CXCL1 and CXCL2) forms a separate node. Similar kind of tree topology for rabbit has also been reported earlier [29]. In summary, this tree outlines the fact that GRO genes have followed specific evolutionary patterns, which might be the outcome of selection pressures.

2.3.2 Structural basis for observed Darwinian selection criterion in GRO chemokines

Prior to selection analysis, GRO genes were analyzed for the occurrence of any gene recombination events as these events can mislead the conclusions obtained from selection analysis. Previous work by Abrantes et al reported the gene conversion events between CC chemokine receptors in various mammalian species, although they are absent in their CC chemokine ligands [30,31]. In order to analyze such gene conversion events in GRO chemokines, GARD and GENECOV analysis was performed. The results evidenced for no statistically significant gene recombination and gene conversion events in GRO genes. These observations are in coherence with the recombination analysis for CC chemokine ligands of CCR5 as described above [30-32].

Selection analysis for GRO genes was performed using maximum likelihood methods of codeml, SLAC, FEL, REL, iFEL, MEME and FUBAR. The selection analysis was individually performed for all the three GRO chemokines (CXCL1, CXCL2, and CXCL3), and also for the complete set of GRO chemokines that comprises of all the sets of these three duplicated Genes. The main aim of performing the analysis for combined set of the GRO chemokines is to assess the correlative nature of duplication and positive selection **(Table 2.1)**. The datasets with individual chemokines returned 3, 5, and 3 sites for CXCL1, CXCL2, and CXCL3 respectively. Some of these sites such as A55, S72 have evolved positively in more than one GRO protein. A total of 17 positively selected sites were observed when the analysis was performed for the entire set of GRO

chemokines. As expected, out of 17 sites, 9 of them are exactly same as obtained from individual analysis. It was presumed that the rest 8 is an outcome of the duplication phenomenon. All the selection analysis results for GRO genes are summarized in **Table 2.1** and **Fig. 2.4**.

The positioning and the importance of these residues were then analyzed with respect to the GRO chemokine structure and function. Among the positively selected residues, four of them (A4, T5, L15, P20) are in the N-terminal domain, four of them (V28, K29, S30, 50A) are in central-core domain and nine of them (A55, I58, K60, E64, S69, D70, K71, S72, N73) are in the C-terminal domain **(Fig. 2.4 A-B)**. It is evident from the data that most of the positively selected sites (10 out of 17 residues) are in the GAG binding surface (C-terminal helix + 3_{10} -helix). Interestingly these residues are mostly spanning on either sides/close proximity of experimentally demonstrated conserved GAG residues (H19, K21 on 3_{10} helix and K61, K65 on C-terminal helix) [2,3]. It has also been observed that the N-terminal region which is crucial for receptor binding and activation is under purifying selection, except the two positively selected residues (L15 and P20). Of these two residues, P20 also overlaps with GAG binding surface. These results indicate that receptor binding domain is majorly showing the signatures of purifying selection, and thus conserved among the GRO chemokines. Indeed such behavior is very much anticipated for the receptor binding domain considering functional constraints of GRO chemokine binding to its conservative CXCR2 receptor. In contrast, the GAG binding domains have evidenced for large number of positive selection sites indicating the evolution of differential GAG binding features/GAG surfaces among GRO chemokines in different species. These results suggest that the two variable and complementary functional segments (receptor binding domain and GAG binding domain) of the GRO proteins have a selection bias in order to balance their structure-function relationship. The other interesting segment of GRO chemokines that showed positive selection is the dimer interface (V28, K29, S30) of the central-core domain. Indeed, such a specific site/segmental selection of GRO chemokines at the GAG binding and their extensions to dimer interface regions, can regulate the functional activity of these chemokines by providing conformational adjustments at dimer interface, and by altering the binding energetics through modulation of positive charge potential on the surface via amino acid substitutions/gene mutations (discussed in later sections).

Table 2.1: *Selection analysis of GRO chemokines using different codon based maximum likelihood methods.*

GRO Genes	No. of species	*lnL*M7	*lnL*M8	2∆lnL	PAML	SLAC[b]	FEL[c]	REL[f]	iFEL[d]	MEME[e]	FUBAR[g]	Total no. of sites
CXCL1	37	-2994.58	-2994.24	0.54	NA	20, 72	20, 72	4, 15, 20, 36, 41, 45, 69, 70, 71, 72	72	20, 69, 72	20, 72	20, 69, 72
CXCL2	22	-1654	-1649.52	8.97*	5, 15, 55, 64	-	4	3, 4, 5, 13, 15, 20, 22, 26, 45, 46, 55, 62, 64, 65, 66, 68, 72, 73	5, 36	-	5, 64	4, 5, 15, 55, 64
CXCL3	24	-1913.98	-1909.55	8.85*	5, 50, 55, 72	-	50	3, 15, 22, 33, 50, 55, 58, 60, 63, 64, 65, 66, 70, 72	46, 55	50, 73	50	50, 55, 72
Combined dataset	83	-5007.58	-5003.71	7.7*	5, 15, 20, 30, 50, 55, 64, 69, 70, 71, 72	4, 15, 20, 36, 70, 72	4, 15, 20 ,50, 70,72	NA	4, 5, 46, 55,72	4, 15, 20, 28, 29, 30, 58, 60, 69, 71, 72, 73	4, 15, 20, 21, 28, 29, 30, 36, 43, 58, 60, 64, 71, 72, 73	4, 5, 15, 20, 28, 29, 30, 50, 55,58, 60, 64, 69, 70, 71, 72, 73

[a]*Amino acids identified by more than one method are underlined.*
[b,c,d,e] *Amino acids with significance values < 0.1*
[f]*Amino acids with Bayes factor > 95*
[g]*Amino acids with significance values >= 0.9*
**P<0.05*

The relationship among these positively selected residues was further analyzed in order to assess their role in coevolution. In coevolution/compensatory evolution, when one amino acid evolved at a particular position, other amino acid evolves at different position simultaneously as a coevolving pair. The coevolving nature of all the positively selected residues was analyzed **(Table 2.2, Fig. 2.4 C)**. These results demonstrated that several of these amino acids are coevolved and contribute to the structural stability through side chain networks via different tertiary interactions.

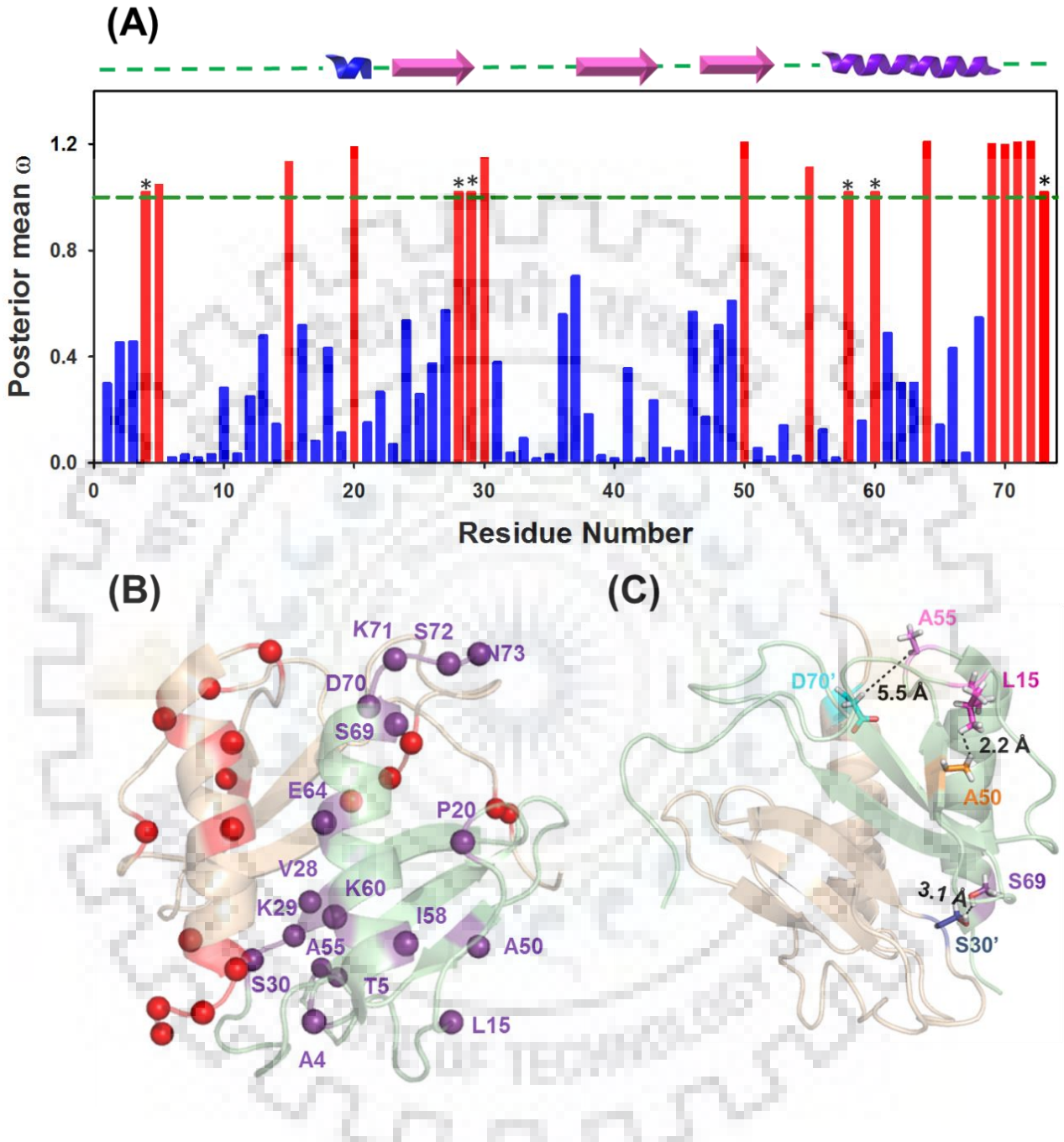


Figure 2.4: (A) *Posterior mean ω at each amino acid site across the GRO genes; Blue-purifying selection, Red-positive selection, and horizontal line represent cutoff value for positive selection. '*' mark indicate the sites detected as positively selected by methods of DATAMONKEY server.* **(B)** *Positively selected residues are marked as spheres on both the monomeric units of HCXCL1 homodimer.* **(C)** *Significant tertiary interactions observed among the residues forming coevolved pairs are shown on HCXCL1 structure.*

Table 2.2: *Calculated average mutual information (AMI) scores for the coevolving positively selected sites. Number of stars represents the significance the coevolving interaction.*

Coevolving amino acid pairs	AMI Score
5T----30S	9.43 (**)
5T----55A	7.29 (*)
30S----20P	6.91 (*)
30S----29K	10.36 (***)
30S----55A	11.56 (***)
69S----70D	9.24 (**)
69S----71K	10.35 (***)
69S----72S	10.91 (***)
71K----72K	10.4 (***)

For example, coevolving pair L15-A50 is involved in hydrophobic interaction. Similarly, L55-D70' and S69-S30' contributes to the inter-subunit interactions **(Fig. 2.4 C)**. It is worth noting that, sometimes coevolved residues are a part of the larger interacting networks. In such cases, a direct interaction with a particular site (positively selected partner) is not evident from distance measurements as noticed for T5. Considering the conserved nature of the receptor domain, and interesting evolutionary patterns of GAG binding domain of GRO chemokines, the evolutionary analysis and discussion is more oriented towards understanding the species specific differential GAG binding characteristics.

2.3.3. Duplication of GRO genes followed concerted evolution

Different models have been proposed for the evolution of multigene families that includes divergent evolution, concerted evolution, and birth and death evolution. In divergent evolution, Genes obtained after being duplicated from their parental genes, acquires new gene functions independently. Whereas in concerted evolution, all the members of a gene family evolve in a concerted/coincidental manner [33,34]. For concerted evolution, dS should be similar to or slightly higher than dN, irrespective of purifying selection. Moreover, if birth-and-death evolution is acting on a gene, then the rate of silent mutation should be extensive, leading dS to be many folds higher than dN [35].

To identify the evolutionary method of GRO genes, the pairwise dS and dN values of GRO genes for different species were compared. It was found that, in each species the dS values were similar or slightly higher than dN values indicating the concerted mode of evolution **(Table 2.3)**. Indeed, concerted evolution has been followed by many genes including ribosomal RNA genes and histone genes that codes for the large quantities of proteins with similar functions and are crucial for the endurance of the organism [36,37].

Table 2.3: *Pair wise values of dS (below diagonal) and dN (above diagonal) observed between GRO sequences (CXCL1/CXCL2/CXCL3 denoted as 1/2/3) from H-human, O-orangutan, S-horse, B-cow, M-mouse and R-rat.*

	H1	H2	H3	O1	O2	O3	S1	S2	S3	B1	B2	B3	M1	M2	M3	R1	R2	R3
H1		0.06	0.09	0.01	0.05	0.08	0.16	0.16	0.15	0.16	0.16	0.14	0.22	0.20	0.26	0.19	0.25	0.29
H2	0.08		0.08	0.06	0.01	0.07	0.11	0.12	0.12	0.13	0.13	0.11	0.21	0.19	0.23	0.18	0.22	0.27
H3	0.15	0.15		0.09	0.08	0.01	0.12	0.12	0.11	0.12	0.12	0.10	0.24	0.18	0.27	0.21	0.21	0.29
O1	0.04	0.08	0.14		0.05	0.08	0.16	0.16	0.15	0.14	0.14	0.14	0.24	0.20	0.25	0.21	0.24	0.28
O2	0.08	0.02	0.06	0.08		0.06	0.12	0.13	0.13	0.13	0.13	0.12	0.21	0.18	0.23	0.19	0.22	0.28
O3	0.16	0.10	0.06	0.11	0.08		0.11	0.12	0.10	0.11	0.11	0.09	0.23	0.17	0.26	0.20	0.21	0.28
S1	0.28	0.41	0.41	0.29	0.35	0.34		0.04	0.05	0.11	0.11	0.07	0.20	0.20	0.25	0.17	0.21	0.26
S2	0.26	0.32	0.32	0.26	0.23	0.26	0.06		0.06	0.10	0.10	0.08	0.22	0.18	0.24	0.17	0.19	0.24
S3	0.33	0.42	0.42	0.27	0.30	0.36	0.14	0.12		0.10	0.10	0.08	0.24	0.22	0.30	0.21	0.24	0.31
B1	0.46	0.39	0.46	0.43	0.36	0.46	0.36	0.28	0.40		0.00	0.07	0.26	0.21	0.28	0.22	0.23	0.28
B2	0.42	0.36	0.42	0.39	0.33	0.42	0.33	0.25	0.36	0.02		0.07	0.26	0.21	0.28	0.22	0.23	0.28
B3	0.51	0.48	0.55	0.56	0.48	0.55	0.38	0.29	0.47	0.33	0.30		0.20	0.19	0.25	0.17	0.20	0.27
M1	0.52	0.53	0.64	0.55	0.53	0.56	0.54	0.39	0.48	0.49	0.52	0.59		0.25	0.29	0.04	0.25	0.27
M2	0.59	0.61	0.70	0.64	0.57	0.61	0.59	0.51	0.53	0.60	0.65	0.72	0.40		0.13	0.22	0.06	0.15
M3	0.70	0.69	0.67	0.66	0.68	0.59	0.54	0.55	0.45	0.42	0.56	0.72	0.51	0.24		0.24	0.13	0.09
R1	0.51	0.62	0.56	0.54	0.57	0.49	0.54	0.43	0.52	0.70	0.65	0.72	0.20	0.46	0.58		0.22	0.29
R2	0.72	0.67	0.79	0.78	0.65	0.69	0.63	0.55	0.65	0.64	0.69	0.71	0.52	0.23	0.24	0.53		0.16
R3	0.96	0.88	0.90	1.04	0.68	0.79	0.65	0.66	0.71	0.76	0.82	0.84	0.54	0.36	0.16	0.68	0.20	

2.3.4. Substitution rate analysis of GRO proteins

Accumulation of mutations in the coding sequences of the duplicated genes greatly influences their functions. They include differential expression patterns, and various interactions with biomolecular binding partners. Selection analysis along with the calculated amino acid

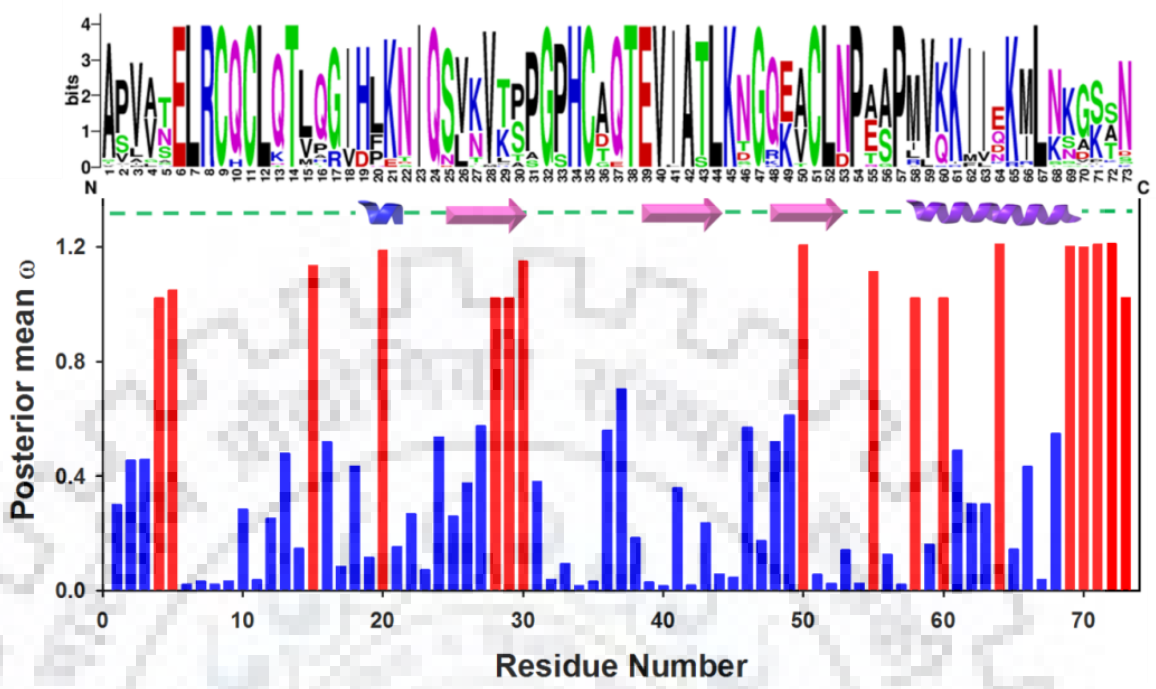


Figure 2.5: *Extent of conservation of GRO sequences among different species calculated using ConSurf server* [22] *along with the sequence logo created by the WebLogo program* [23]. *The positively selected residues are marked with red bars.*

conservation scores suggested that all the positively selected sites exhibit a conservation score below 70 % and several of the nucleotides/amino acids that correspond to the purifying/neutral selection do vary considerably **(Fig. 2.5)**. Therefore, in order to throw light on the rate and nature of nucleotide substitutions resulted after duplication events, the substitution rates were calculated **(Fig. 2.6)**. It was observed that rodent's genes are more prone to substitution in comparison to other species, when substitution rates were calculated for the full length sequence of all the species **(Fig. 2.6A)**. Further, to determine the segments of the protein that has major contribution to the high substitution rates, the protein was divided into three domains: N-terminal domain, central domain, and C-terminal domain **(Fig. 2.1A)**.

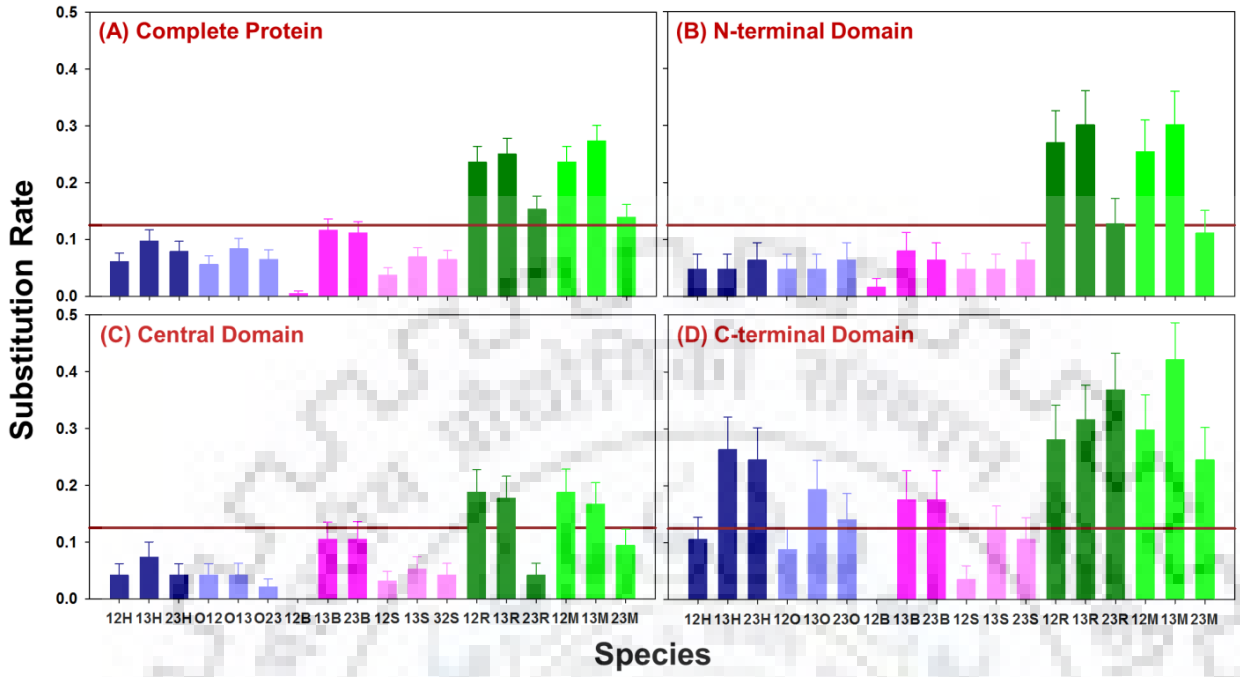


Figure 2.6: *Substitution rates for* **(A)** *complete GRO sequences, and individual domains namely;* **(B)** *N-terminal,* **(C)** *Central, and* **(D)** *C-terminal domain. H-human, O-Orangutan, S-Horse, B-Cow, R-Rat, M-Mouse and 1/2/3 represent CXCL1/CXCL2/CXCL3. The brown horizontal line represents the lower threshold of substitution rate.*

Substitution rates were calculated for each of the domain among all the species. It was observed that in primates and laurasiatherians, C-terminal domain alone showed the high substitution rates, suggesting that the rest of the protein is under stringent purifying selection **(Fig. 2.6B-C)**. However, in contrast to the higher order mammals, both N-terminal and C-terminal domains showed high substitution rates in the rodents. Moreover, the C-terminal domain has highest propensity of substitution changes compared to the N-terminal counterpart **(Fig. 2.6D)**. Further, the comparative analysis of the substitution rates of alone C-terminal domain of the rodents with that of the same in primates and laurasiatherians suggested that rodents accumulated higher substitution rates in the C-terminal domain.

To further support conclusions on the variable substitution pattern, the pair wise values of ω for different parts of the GRO proteins among different species were compared **(Fig. 2.7)**. In primates, higher ω values were observed in case of C-terminus as compared to middle and N-

terminus (C-terminal domain > N-terminal domain > Central domain). In Laurasiatherians, a slightly different trend in ω values was observed, i.e., C-terminal domain > N-terminal domain = Central domain. Rodents showed a distinctive pattern of ω values in which C-terminal domain = N-terminal domain > central domain. All these results clearly point towards the species specific evolutionary changes regulated by variable selection pressures on different parts of the protein for a functional significance/advantage.

2.3.5. Sequence and surface properties of GRO proteins

Selection pressure analysis at domain level of GRO proteins prompted for further delineation of the molecular level details and sources of differential ω value patterns across different species. In order to dig out the differential evolution characteristics of GRO proteins across the species, the variations in their sequences at both nucleotide and amino acid level were simultaneously analyzed **(Fig. 2.8)**. Very small number and specifically regulated nucleotide and amino acid changes were observed across GRO chemokines in primates and laurasiatherians. Moreover, the changes in these species are majorly through a facile single nucleotide alterations, in order to guide the duplicated sibling for a functional advantage via minimal perturbation. In contrast to this, rodent genes exhibited large number of changes with multitude of nucleotide alterations accompanying for amino acid makeovers. It is worth noting that in case of primates, the amino acid alterations are on the helical surface in GRO genes i.e., essentially confined within the C-terminal α-helix. Whereas in rodents, apart from the N-terminal, these changes have spanned across both 3_{10} helix along with C-terminal α-helix that constitute the GAG binding surface of their positively charged residues. Moreover, the observed gene mutation(s) in the GAG binding surfaces specifically corresponds to the change(s) in amino acids that result in the charge reversal/charge neutralization of protein surface. These observations incriminate that, promiscuity of GRO genes under evolutionary pressure is focused towards modulating the GAG binding surfaces via minimal gene alterations through charge distribution.

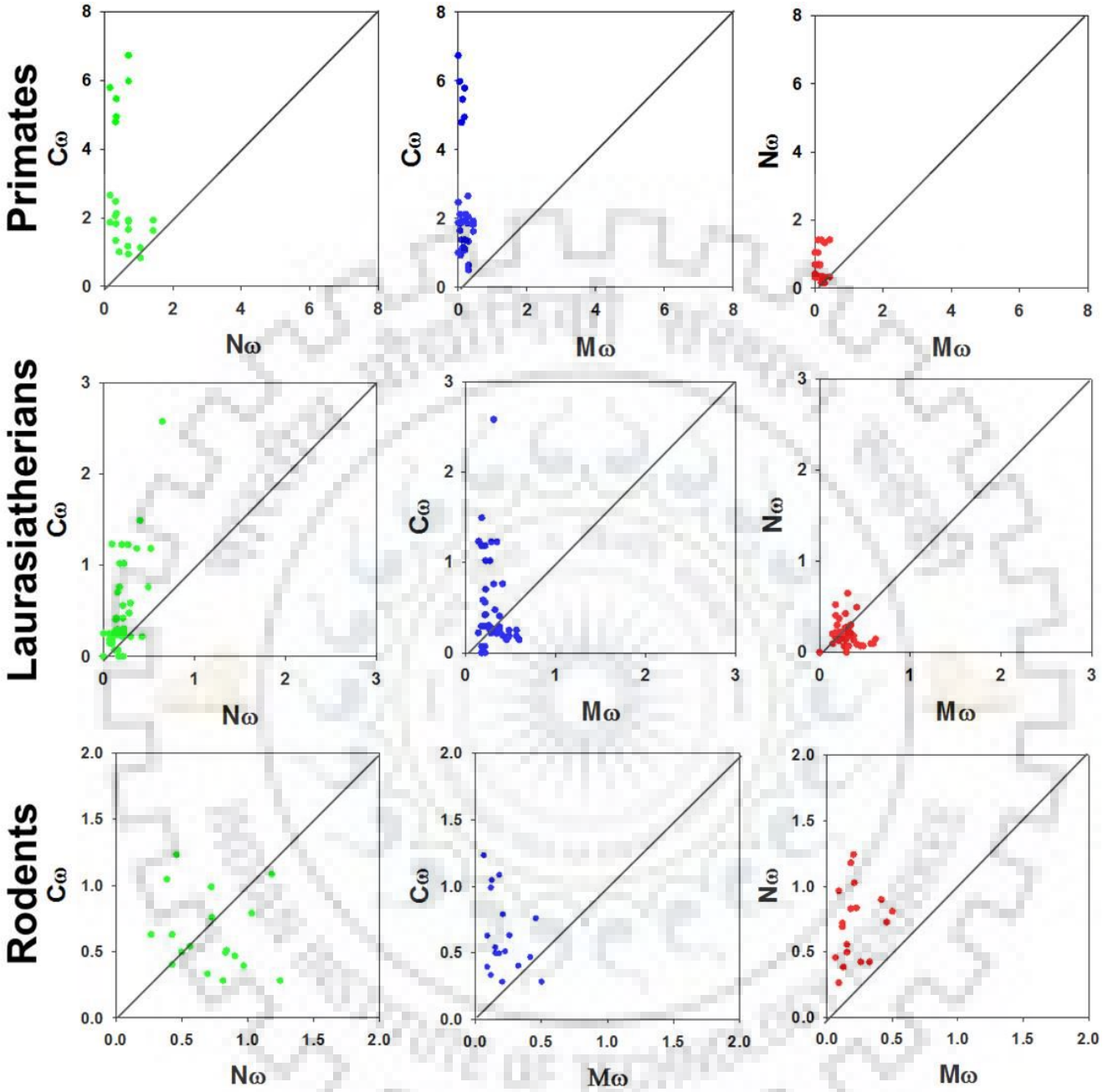


Figure 2.7: *Pair wise comparison of omega values for different domains of GRO genes in different families. Cω, Mω and Nω represents the ω values for C-terminal, central and N-terminal domains respectively.*

Figure 2.8: *Comparative nucleotide/amino acid analysis of GRO chemokines; H-human, S-horse, M-murine and 1/2/3 represent CXCL1/CXCL2/CXCL3. Alterations in the nucleotide/amino acid sequences are highlighted in red.*

To delineate the surface architectures and electrostatic potentials that can plausibly contribute to differential GAG binding, the dimeric structures of murine, horse, and human GRO chemokines were modeled **(Fig. 2.9)**. These structural models furnished a pictorial view of distinct GAG binding surfaces that are generated by virtue of amino acid changes. Analysis on human GRO proteins (as dimers), yielded complex electrostatic surfaces and evoked the feasibility of multiple modes of GAG binding yielded complex electrostatic surfaces and evoked the feasibility of multiple modes of GAG binding **(Fig. 2.9, H-upper panel)**. It is evident from the electrostatic potential pattern that the human GRO genes will have an alternate GAG binding surface comprising of the positive residues (R8, K29, R48, K49; according to HCXCL1) **(Fig. 2.1A)**.

In line with evolutionary theory, a recent study on HCXCL1 confirmed the presence of two non-overlapping GAG-binding domains in HCXCL1 [3]. Interestingly, due to surface charge alterations, HCXCL2 has evolved with a contrasting 90^{o} rotated positive surface on the β-sheet/dimer interface as compared to HCXCL1 and HCXCL3 **(Fig. 2.9, H-upper panel)**. Such a surface of HCXCL3 is due to the presence of K27 along with R/K29 at the dimer interface and

lack of R48 at the start of β_3-strand and is a resultant of single nucleotide alteration at these two sites on the β-surface of human GRO proteins. Moreover, these changes resulted in specific positioning and shuttling of the residues within Arg and Lys (K29, R48 in HCXCL1 vs R29, K48 in HCXCL3) to fine tune their GAG binding interactions/affinities. Such a GAG binding interface at the β-domain surface has also been reported for homeostatic chemokine CXCL12 [38].

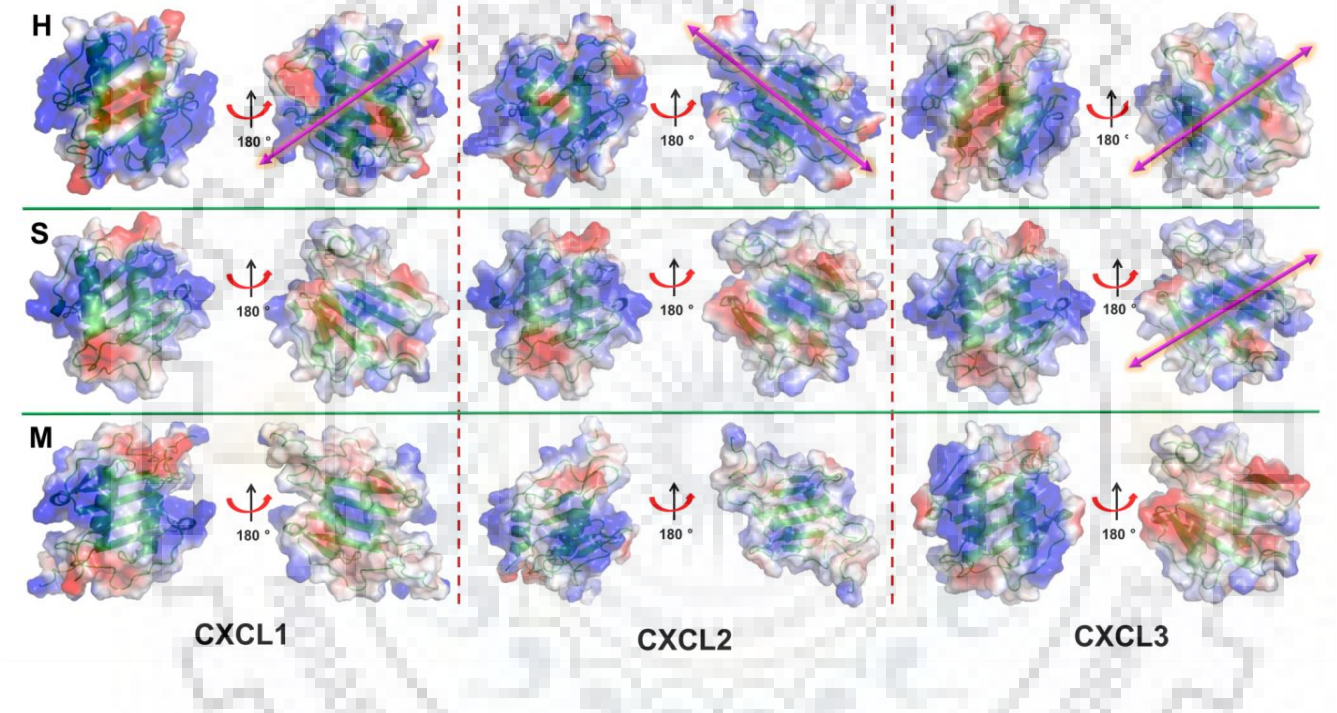


Figure 2.9: *Electrostatic surface potential maps for the dimeric GRO proteins. The upper, middle and lower panel represent the H-human, S-horse, M-murine chemokines respectively. Both α-helical and β-sheet surfaces are presented. Pink arrows represent the plausible mode of GAG binding on β-sheet surfaces. The vacuum electrostatics was generated using PyMOL molecular graphics system[28].*

Similar analysis with horse proteins yielded surfaces that are different from the primates. In horse CXCL1 (SCXCL1), the positive surface is confined at the 3_{10} helices on the helical side with a small positive patch at the beta surface of dimer interface **(Fig. 2.9, S-middle panel)**. The positive charges on both the helical and the beta surfaces enhanced considerably in SCXCL3 as compared to SCXCL1 and SCXCL2 **(Fig. 2.9, S-middle panel)**. This can be attributed to the

single nucleotide alterations leading to (a) charge reversal by the change of E49 to K49, (b) charge neutralization by the change of E55 to A55, and (c) contribution of enhanced positive charge with a replacement of K65 to R65 in SCXCL3. Such an enhanced positive electrostatic potential on both surfaces indicates the plausible multiple GAG binding modes of SCXCL3 as compared to SCXCL1/2.

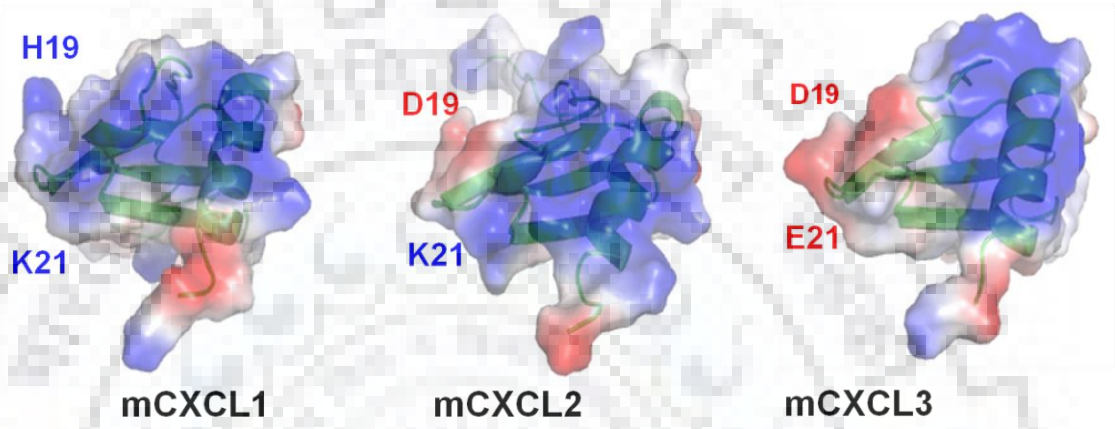


Figure 2.10: *Electrostatic surface potential maps for murine GRO proteins in monomeric form.*

In case of rodents, the mechanism of differential GAG surface formation is different from those of the primates and laurasiatherians. In MCXCL1, 3_{10} -helix possess positive charge due to presence of positively charged residues H19 and K21 **(Fig. 2.9, M-lower panel**, and **Fig. 2.10)**. Whereas in MCXCL2, H20 is replaced by negatively charged D19 and in MCXCL3, whole positive charge at the 3_{10} helix is transformed to negative charge as a consequence of D19 and E21 **(Fig. 2.9, M-lower panel)**. Such site specific charge reversal features have resulted in varied positive surfaces in the murine CXCL1-3, although no positive potentials were observed on the beta surfaces of these proteins **(Fig. 2.9 M-lower panel)**. These positive surface potentials constitute single GAG binding motif on the helical surface in murine GRO proteins, and indeed previous NMR experiments on MCXCL1-trapped dimer demonstrated that GAG binds orthogonal to the inter-helical axis and the residues H20, K22 on 3_{10} helix and K60 and K64 on the C-terminal helix are crucial for GAG binding [2].

Considering all the above observed electrostatic surface variations on the mammalian GRO proteins **(Fig. 2.8 & Fig. 2.9)**, it is summarized that the β-sheet binding surface and its relative

surface orientation(s) are confined to the primates and laurasiatherians and no such surface is evident in the rodents. Further, analysis suggested that the electrostatic surfaces have evolved specifically among the GRO genes in a given organism, in order to serve the needs of particular species as per the evolutionary pressure, although the helical surface of GAG binding resembles the primordial mode of GAG binding in all the species. Indeed, selection analysis supported this theme as the residues H19, K21, on 3_{10} helix and K61 and K65 on the C-terminal helix that are crucial for GAG binding are on a tight purifying selection for all the species.

2.4 Discussion

2.4.1 Mechanistic insights into evolutionary characteristics of GRO chemokines

Immunity and defense related genes evolve rapidly under positive selection pressures accumulating amino acid changes more rapidly than other genes. Chemokines listed as one of the 8 most rapidly changing proteins and domains [39,40]. The reason for continuous evolvability of inflammatory chemokines is intimately related to their functionality in host defense processes, as viruses copy the host endogenous chemokines, and target the host machinery. In order to fight against such evolving pathogens, the host genes need to undergo smart changes and also expand themselves to maintain their diversity.

The work in this chapter is an effort to furnish mechanistic insights into the evolutionary characteristics of GRO chemokines. The phylogenetic analysis provided an overview of the fact that GRO genes have evolved in species specific patterns. The observed phylogenetic profile for GRO genes is very much consistent with the earlier literature available on CXC chemokines [29,41,42]. Indeed, such an enchanting species specific evolution of GRO genes is an outcome of the mechanism of duplication events that has been precisely reflected in the genomic arrangements of different species including primates, laurasiatherians, and rodents [42]. In rodents, all GRO proteins are present in same transcriptional orientation in contrast to primates where CXCL2 and CXCL3 are placed in an opposite direction to CXCL1, as primate GRO genes evolution is an outcome of inverse duplication as compared to other species [43]. Such a gene positioning via duplication and species specific nucleotide alterations under evolutionary pressure confer multiple layers of regulation to chemokine functions in terms of differential transcriptional regulation,

different cellular territories of distribution, homo/hetero dimerization, and receptor/GAG recognition to form cell/chemokine specific chemotactic gradients/cell signaling events. Indeed, experimental reports on human CXCL1, CXCL2, and CXCL3 suggested that the expression of these three highly similar genes is differentially regulated in a cell specific/signal specific manner, and binds to CXC receptors with varied affinities [4,5,29].

Site specific selection analysis of duplicated GRO chemokines confirmed that the genes underwent majorly through purifying/negative selection with few positively selected sites, several of them are coevolving and majority of them are lying in the GAG binding domain/dimerization region. In general, majority of the genes evolve under negative selection. For example, fork head gene family has evolved as a result of gene duplication, rapid differentiation and subsequent fixation of amino acid changes through negative selection [44]. Similarly, the post duplication charge evolution of surface charge in phosphoglucose isomerase was not driven by strong selection on individual amino acid sites but by the weak selection on large number of amino acid sites and consequently by steady directional/purifying selection on overall structural properties of the protein, which are derived from many modifiable sites [45]. In the case of GRO genes also, the modification of the electrostatic surfaces are not in line with the principles of strong positive selection. They evolved majorly under purifying selection with an exception of C-terminal domain. Such domain specific evolutionary patterns are seen in immune signaling protein TRAF3 interacting protein 2. In this protein, the N-terminal domain that is more disordered has been subjected to positive selection, and a purifying selection was observed for the C-terminal domain which is contributing to the core structure of the protein [46]. For GRO chemokines, the selection, substitution rates, and nucleotide comparison studies established that significant degree of positive selection/alterations are concentrated in the C-terminal/GAG binding domain including residues involved in the dimerization contacts. However, no positive selection has been noticed in the structural part of the protein suggesting the conserved nature of the structural fold for its chemotactic functioning.

This analysis also established that such evolutionary changes resulted in multiple species specific pathways for differential GAG binding such as regulated electrostatic surface in the 3_{10} helix in rodents and generation of novel β-sheet surface in primates and some of the GRO proteins

in laurasiatherians, although the helical surface binding appears to be the primordial mode of GAG binding for GRO chemokines. Evolutionary analysis on formyl peptide receptors across mammals evidenced for similar sort of differential electrostatic surfaces through amino acid alterations [47]. Essentially, chemokine immobilization through GAGs facilitate the formation of haptotactic gradient, and adds another layer of specificity and control to the cell migration beyond the receptor. These evolved species specific and gene specific GRO proteins can also have differential half-life and susceptibility to protease degradation upon GAG binding.

Further, these species specific evolutionary events are also credited for gene inactivation and partial/complete deletions leading to pseudogenes (non-functional), like genes (novel functions) and extinct genes (erased) [39,48-51]. Like genes for CXCL1 and CXCL3 (CXCL1L/CXCL3L) were reported in *Macaca fascicularis* [48]. In both the cases, although the contributing GAG residues are similar, the C-terminal region that is crucial for monomer-dimer equilibrium is truncated thus contributing to differential populations of GAG induced dimerization/oligomerization. On a similar note, a pseudogene (CXCL1p) found in human and chimpanzee in the inversely duplicated segment of chromosome 4, which is yet to be processed completely for complete functionality. Currently CXCL1p gene contains only two exons with one intron in-between, accompanied by downstream deletion resulting in a sequence that is devoid of C-terminal helix [48]. Sequence analysis identified that in CXCL1p, residue R9 that is essential for GAG/receptor binding has been mutated to P9 thus switching off the functionality of this gene at this stage.

Such gene alterations are not only confined to the GRO family chemokines, several other CXC chemokines also experienced these evolutionary pressures for functional advantages. For example, CXCL4L1, a homologue of Platelet factor 4 (CXCL4) is a resultant of duplication of CXCL4 gene present in humans and chimpanzees, which differs in only three amino acids (P58L, K66E, and L67H) as compared to CXCL4 and exhibits potent anti-angiogenic and anti-tumorous properties. Recent structural studies have shown that C-terminal helix of CXCL4L1 adopts an open conformation due to L67H mutation. These amino acid changes coupled with the conformational transition in the C-terminal helix is responsible for the lowering of GAG binding affinities of CXCL4L1 as compared to CXCL4 [52].

Another essential feature is, during the course of evolution, these genes have also acquired variations due to point mutations. Different types of polymorphisms like single nucleotide polymorphisms, insertion and deletions of specific nucleotides, alternating splicing that may be associated with several diseases or sometimes may be beneficial [39]. CXCL2 which is an essential mediator in lung protection but polymorphism of a short tandem repeat (AC)n at -665 position in the promoter region altered the promoter's activity, consequently heightened the expression of CXCL2 thus contributing to severe sepsis [53]. In such pathological cases, redundancy/promiscuity of these duplicated inflammatory chemokines play a key role due to an overlap in major functions carried out by them; so that any defect in one chemokine can be easily resolved by an alternate family member to safeguard the cell from immune insults.

2.5 Conclusions

In summary, the current chapter explored the evolutionary history of GRO genes across the diverse range of mammals. These molecular evolutionary studies on GRO proteins threw light on the underlying principles responsible for their variable evolutionary patterns, positive selection of the protein segments, substitution rates and their lineage to GAG binding properties thus bridging our structural awareness with evolutionary programming and functional variance. Phylogenetic analysis of GRO chemokines across mammalian species showed a trend of species specific evolution pattern. Selection analysis revealed that although the GRO genes underwent purifying selection, but encompasses 17 positively selected sites that are majorly lying in the GAG binding segments. Some of these positively selected sites are under compensatory evolution as depicted by the coevolutionary analysis. Further, GRO genes have evolved coincidentally, implicating that they followed the concerted mode of evolution. Furthermore, substitution rates and comparative analysis of pairwise ω values among different species implied that different parts of the GRO proteins experienced different selection pressures in species dependent manner for their functional benefits. Moreover, such alterations in their nucleotide/amino acid sequences resulted in the species specific evolution of electrostatic surfaces that are defining different geometries for GAG binding. Future comparative experimental studies involving protein-GAG interactions of GRO proteins from multiple species are imperative to decipher the regulatory role of multiple GAG binding surfaces/orientations during chemotactic gradients and hence neutrophil trafficking. Such a

detailed knowledge is applicable to all the protein/chemokine families that demand a detailed lineage of evolution-structure-function relationships.

2.6 References

1. Zlotnik A, Yoshie O. The chemokine superfamily revisited. Immunity. 36(5), 705-716 (2012).

2. Poluri KM, Joseph PR, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. J. Biol. Chem. 288(35), 25143-25153 (2013).

3. Sepuru KM, Rajarathnam K. CXCL1/MGSA is a novel glycosaminoglycan (GAG)-binding chemokine: structural evidence for two distinct non-overlapping binding domains. J Biol. Chem. (2015).

4. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, Rousseau S, Eidelman DH, Hamid Q. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. J Immunol. 191(5), 2731-2741 (2013).

5. Luttichau HR. The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist. J Biol. Chem. 285(12), 9137-9146 (2010).

6. Sajadi SM, Khoramdelazad H, Hassanshahi G, Rafatpanah H, Hosseini J, Mahmoodi M, Arababadi MK, Derakhshan R, Hasheminasabzavareh R, Hosseini-Zijoud SM, Ahmadi Z. Plasma levels of CXCL1 (GRO-alpha) and CXCL10 (IP-10) are elevated in type 2 diabetic patients: evidence for the involvement of inflammation and angiogenesis/angiostasis in this disease state. Clin. Lab. 59(1-2), 133-137 (2013).

7. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539- (2011).

8. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30(12), 2725-2729 (2013).

9. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol Evol. 23(10), 1891-1901 (2006).

10. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 26(19), 2455-2457 (2010).

11. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. Virology. 265(2), 218-225 (1999).

12. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24(8), 1586-1591 (2007).

13. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS. Genet. 8(7), e1002764- (2012).

14. Poon AF, Frost SD, Pond SL. Detecting signatures of selection from DNA sequences using Datamonkey. Methods Mol. Biol. 537, 163-183 (2009).

15. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol. Biol Evol. 30(5), 1196-1205 (2013).

16. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22(4), 1107-1118 (2005).

17. Pinheiro A, Woof JM, bi-Rached L, Parham P, Esteves PJ. Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. PLoS. One. 8(9), e73934- (2013).

18. Pinheiro A, Woof JM, Almeida T, Abrantes J, Alves PC, Gortazar C, Esteves PJ. Leporid immunoglobulin G shows evidence of strong selective pressure on the hinge and CH3 domains. Open. Biol. 4(9), 140088- (2014).

19. Neves F, Abrantes J, Esteves PJ. Evolution of CCL11: genetic characterization in lagomorphs and evidence of positive and purifying selection in mammals. Innate. Immun. 22(5), 336-343 (2016).

20. Lemos de MA, McFadden G, Esteves PJ. Evolution of viral sensing RIG-I-like receptor genes in Leporidae genera Oryctolagus, Sylvilagus, and Lepus. Immunogenetics. 66(1), 43-52 (2014).

21. Lemos de MA, McFadden G, Esteves PJ. Positive evolutionary selection on the RIG-I-like receptor genes in mammals. PLoS. One. 8(11), e81864- (2013).

22. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44(W1), W344-W350 (2016).

23. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 14(6), 1188-1190 (2004).

24. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3(5), 418-426 (1986).

25. Steinway SN, Dannenfelser R, Laucius CD, Hayes JE, Nayak S. JCoDA: a tool for detecting evolutionary selection. BMC. Bioinformatics. 11, 284- (2010).

26. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino BC. MISTIC: Mutual information server to infer coevolution. Nucleic Acids Res. 41(Web Server issue), W8-14 (2013).

27. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 18(15), 2714-2723 (1997).

28. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.4.1.

29. Modi WS, Yoshimura T. Isolation of novel GRO genes and a phylogenetic analysis of the CXC chemokine subfamily in mammals. Mol. Biol. Evol. 16(2), 180-193 (1999).

30. Abrantes J, Carmo C, Matthee C, Yamada F, Loo W, and Esteves P. A shared unusual genetic change at the chemokine receptor type 5 between Oryctolagus, Bunolagus and Pentalagus. Conservation Genetics12, 325- (2011)

31. de Matos AL, Lanning DK, Esteves PJ. Genetic characterization of CCL3, CCL4 and CCL5 in leporid genera Oryctolagus, Sylvilagus and Lepus. Int. J Immunogenet. 41(2), 154-158 (2014).

32. Carmo CR, Esteves PJ, Ferrand N, van der LW. Genetic variation at chemokine receptor CCR5 in leporids: alteration at the 2nd extracellular domain by gene conversion with CCR2 in Oryctolagus, but not in Sylvilagus and Lepus species. Immunogenetics. 58(5-6), 494-501 (2006).

33. Smith GP. Unequal crossover and the evolution of multigene families. Cold Spring Harb. Symp. Quant. Biol. 38, 507-513 (1974).

34. Smith GP, Hood L, Fitch WM. Antibody diversity. Annu. Rev. Biochem. 40, 969-1012 (1971).

35. Rooney AP, Piontkivska H, Nei M. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol. Biol. Evol. 19(1), 68-75 (2002).

36. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121-152 (2005).

37. Matsuo Y, Yamazaki T. Nucleotide variation and divergence in the histone multigene family in Drosophila melanogaster. Genetics. 122(1), 87-97 (1989).

38. Ziarek JJ, Veldkamp CT, Zhang F, Murray NJ, Kartz GA, Liang X, Su J, Baker JE, Linhardt RJ, Volkman BF. Heparin oligosaccharides inhibit chemokine (CXC motif) ligand 12 (CXCL12) cardioprotection by binding orthogonal to the dimerization interface, promoting oligomerization, and competing with the chemokine (CXC motif) receptor 4 (CXCR4) N terminus. J Biol. Chem. 288(1), 737-746 (2013).

39. Colobran R, Pujol-Borrell R, Armengol MP, Juan M. The chemokine network. II. On how polymorphisms and alternative splicing increase the number of molecular species and configure intricate patterns of disease susceptibility. Clin. Exp. Immunol. 150(1), 1-12 (2007).

40. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ. Defining the origins and evolution of the chemokine/chemokine receptor system. J Immunol. 176(1), 401-415 (2006).

41. Widdison S, Coffey TJ. Cattle and chemokines: evidence for species-specific evolution of the bovine chemokine system. Anim Genet. 42(4), 341-353 (2011).

42. Nomiyama H, Osada N, Yoshie O. The evolution of mammalian chemokine genes. Cytokine Growth Factor Rev. 21(4), 253-262 (2010).

43. Shibata K, Nomiyama H, Yoshie O, Tanase S. Genome diversification mechanism of rodent and Lagomorpha chemokine genes. Biomed. Res. Int. 2013, 856265- (2013).

44. Fetterman CD, Rannala B, Walter MA. Identification and analysis of evolutionary selection pressures acting at the molecular level in five forkhead subfamilies. BMC. Evol. Biol. 8, 261- (2008).

45. Sato Y, Nishida M. Post-duplication charge evolution of phosphoglucose isomerases in teleost fishes through weak selection on many amino acid sites. BMC. Evol. Biol. 7, 204- (2007).

46. Wu B, Gong J, Yuan S, Zhang Y, Wei T. Patterns of evolutionary selection pressure in the immune signaling protein TRAF3IP2 in mammals. Gene. 531(2), 403-410 (2013).

47. Muto Y, Guindon S, Umemura T, Kohidai L, Ueda H. Adaptive evolution of formyl peptide receptors in mammals. J Mol. Evol. 80(2), 130-141 (2015).

48. Nomiyama H, Otsuka-Ono K, Miura R, Osada N, Terao K, Yoshie O, Kusuda J. Identification of a novel CXCL1-like chemokine gene in macaques and its inactivation in hominids. J Interferon Cytokine Res. 27(1), 32-37 (2007).

49. Struyf S, Burdick MD, Peeters E, Van den BK, Dillen C, Proost P, Van DJ, Strieter RM. Platelet factor-4 variant chemokine CXCL4L1 inhibits melanoma and lung carcinoma growth and metastasis by preventing angiogenesis. Cancer Res. 67(12), 5940-5948 (2007).

50. Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome Biol. 7(12), 243- (2006).

51. Dubrac A, Quemener C, Lacazette E, Lopez F, Zanibellato C, Wu WG, Bikfalvi A, Prats H. Functional divergence between 2 chemokines is conferred by single amino acid change. Blood. 116(22), 4703-4711 (2010).

52. Kuo JH, Chen YP, Liu JS, Dubrac A, Quemener C, Prats H, Bikfalvi A, Wu WG, Sue SC. Alternative C-terminal helix orientation alters chemokine function: structure of the anti-angiogenic chemokine, CXCL4L1. J Biol. Chem. 288(19), 13522-13533 (2013).

53. Flores C, Maca-Meyer N, Perez-Mendez L, Sanguesa R, Espinosa E, Muriel A, Blanco J, Villar J. A CXCL2 tandem repeat promoter polymorphism is associated with susceptibility to severe sepsis in the Spanish population. Genes Immun. 7(2), 141-149 (2006).

Chapter 3: Deciphering the Oligomerization and GAG Binding Characteristics of CXCL1 and CXCL2 Chemokines

Abstract

Chemokines share the fundamental property of oligomerization, they regulate the leukocyte migration via interacting with glycosaminoglycans and G-protein coupled receptors. In the current chapter, the homo-oligomerization and hetero-oligomerization potencies of CXCL1 (GRO α) and CXCL2 (GRO β) were explored. Further, the regulatory role of various synthetic GAGs/GAG mimetics on the oligomerization properties of CXCL1 and CXCL2 was assessed. Studies evidenced for the differential homo oligomerization potentials and heterodimer forming capabilities of CXCL1 and CXCL2. Furthermore, NMR based GAG binding studies suggested that GAGs indeed enhance the oligomerization efficacy of CXCL1 and CXCL2 homo and heterodimers. These results also indicated that pure sulfation of GAGs alone is not sufficient for shifting the CXCL1/CXCL2 chemokine oligomerization equilibrium. The extent/positioning of the sulfation also play a significant role in regulating the chemokine GAG interactions and GAG induced chemokine oligomerization.

3.1 Introduction

Oligomerization is an indispensable regulatory mechanism by which chemokines generate differential/sustainable chemotactic gradients on the endothelial cell surface during the influx of the migrating leukocytes [1]. As mentioned in **Chapter 1**, chemokines commonly adapt two different types of dimers (CXC/CC-type) [2]. Further, several chemokines such as CXCL4, CXCL7, CXCL10, CXCL12 and CCL5 are reported to form tetramers and higher order oligomers by involving both CXC and CC dimeric surfaces [3-7]. Chemokines possessing common structural fold, oligomerizes in similar fashion by exploiting their limited set of dimer interface residues. This augments the feasibility for different chemokines to form hetero-dimeric or higher order hetero-oligomeric structures [8]. Indeed, recent studies have identified that some chemokines pairs including CCL3-CCL4, CXCL4-CCL5, CXCL1-CXCL7, CXCL4-CXCL8, CCL21-CXCL13, CXCL9-CXCL12, CCL2-CCL8 undergo hetero-

oligomerization [8-15]. Furthermore, it has also been shown that the formation of some of these hetero-oligomeric structures exerts functional consequences [9,11,14,16]. This implies that the phenomenon of hetero-oligomerization is specific for some chemokines and adds another layer of mechanism in modulating their in-vivo actions by differentially interacting with the GPCR's and glycosaminoglycans (GAGs) [8].

Additionally, it has been demonstrated that chemokines oligomerization and GAG binding are intimately coupled processes. It is also evident that GAG binding to chemokines induces chemokine oligomerization, increase their local concentrations, and thereby heightens the chemokine mediated cellular processes [17,18]. For example, the non-oligomerizing variants of CCL2/MCP-1, CCL4/MIP-1β and CCL5/RANTES, were not able to recruit the cells in vivo, as was the case with their non-GAG binding variants [19]. Moreover, chemokine-GAG complexes established that chemokine-GAG interactions are highly specific in nature, and also marked various hot spot residues on chemokines that are directly involved in GAG binding [20-22]. On the similar note, it can be presumed that chemokine-GAG interactions can also regulates chemokine hetero-oligomerization. Thus, chemokine homo- and hetero-oligomerization and GAG binding constitute the uppermost synergistic events that are imperative to delineate them at molecular level to understand the chemokine mediated cellular processes.

Studies on NACs revealed the fact that NACs possess intrinsic tendency to form higher oligomerization states (both dimers and tetramers). Despite belonging to same subfamily and sharing high sequence/structural identity, a significant variation in their oligomerization potencies is reported [23]. However, limited knowledge is available regarding the formation of hetero oligomerization within the NACs. In order to shed light on the homo-/hetero-oligomerization potencies of NACs, murine CXCL1 (GRO α) and CXCL2 (GRO β) were chosen, for which the NMR/crystal structures were known [21,22,24]. The current chapter explored CXCL1/CXCL2 oligomerization potencies using biophysical and biochemical experimental analysis. Further, to get insights into the GAG induced chemokine oligomerization, the effect of GAG binding on CXCL1 and CXCL2 homo-/hetero-oligomerization is appraised using different GAGs/GAG mimetics.

3.2 Materials and methods

3.2.1 Cloning of CXCL1 and CXCL2

Full length murine CXCL1 and CXCL2 genes [NCBI Ref seq: NM_008176.3 (CXCL1), NM_009140.2 (CXCL2)] were purchased from Paras industries Ltd (India). The genes encoding the CXCL1/CXCL2 proteins were amplified using appropriate set of forward (FP) and reverse primers (RP) including:

CXCL1_FP:5'-G G T A C C G A A A A C C T G T A T T T T C A G G G A G G T G C T C C G A T T G C T A A C G -3'
CXCL1_RP:5'- C C A T G G T T A C T T C G G A A C G C C C T T C -3'

CXCL2_FP:5'-G G T A C C G A A A A C C T G T A T T T T C A G G G A G C T G T C G T T G C G A G T G A A-3'
CXCL2_RP:5'- C C A T G G T T A G T T T G C T T T G C C T T T G T T C -3'

The following steps were followed to sub-clone the amplified CXCL1/CXCL2 genes into pET32 bacterial expression vector between the Kpn1 and Nco1 restriction sites. The amplified DNA obtained after 30 cycles of PCR (Polymerase chain reaction), was analyzed on 1.5 % agarose gel. The PCR product was cleaned up using PCR clean up kit (Sure Extract, Genetix, India), and subsequently digested by the Kpn1 and Nco1 restriction endonucleases. The pET32 vector was also digested with same pair of Kpn1/Nco1 restriction enzymes. Digestion of amplified gene and vector was carried out for 3 hrs at 37 ºC. Digested products were analyzed on 0.8 % agarose gel and were extracted from the gel using a gel extraction kit (Sure Extract, Genetix, India). Digested gene and vector were ligated using T4 DNA ligase at 16 ºC for 20 hrs. The ligated product at two different volumes (2 µl and 5 µl) was transformed in *E. coli* DH5α competent cells using standard transformation protocol. The transformed cells were plated on LB (Luria-Bertani) plates with 100 µg/ml ampicillin and were incubated at 37 ºC overnight. Around 5-10 colonies were obtained in each plate. Individual colonies were picked up from each plate and inoculated in fresh LB medium containing 100 µg/ml of ampicillin. The cultures were grown in shaker incubator at 37 ºC, with shaking at 200 rpm, for 12-14 hrs. Plasmids were isolated from the cells using Qiagen plasmid DNA kit. The plasmids were checked for the insertion by double digestion using Kpn1/Nco1, and by PCR amplification of the gene, and finally confirmed by DNA sequencing.

3.2.2 Expression and purification of CXCL1 and CXCL2

Plasmids encoding for the corresponding genes were transformed in *E. coli* BL21 (DE3) competent cells. Transformed cells were used to grow overnight seed culture (10 ml), and transferred to 1 L large culture of LB medium or isotopically enriched $^{15}N/^{13}C$ minimal medium containing $^{15}NH_4Cl/^{13}C$-glucose as a sole source of nitrogen and carbon in the presence of 100 μg/ml ampicillin. The cultures were grown at 37 ºC, 220 rpm until its OD reaches 0.6 at 600 nm. Fusion proteins were expressed by inducing the cultures with 0.2 mM isopropyl 1-thio-β-D galactopyranoside (IPTG) and were grown at 20 ºC for 20 hours. To analyze the expression patterns, 1 ml of cells were harvested separately by centrifugation before and after the induction and were resuspended in SDS gel loading buffer containing 100 mM β-mercaptoethanol (BME). The samples were heated at 95 ºC for 20 min and loaded on to a 12 % SDS-PAGE. Cells from the 1L culture were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, pH 8), treated with lysozyme (100 μg/ml) on ice for 1 hour and were lysed by sonication. Cell lysates were centrifuged at 14000 rpm for 60 minutes at 4 ºC and supernatant was separated. After centrifugation, black colored cell debris was observed without the presence of white colored pellet indicating the absence of inclusion bodies, thus suggesting the presence of protein in cytoplasmic/soluble fraction (supernatant). Proteins (CXCL1 and CXCL2) found in cytoplasmic fraction (supernatant) were then purified using Ni-NTA column pre-equilibrated with lysis buffer (20 mM Tris, 500 mM NaCl pH 8). Non-specific proteins were removed by sequential washing using wash buffers 1 and 2 that contains 10 mM and 30 mM imidazole respectively in lysis buffer. Proteins were eluted in the same lysis buffer using 400 mM imidazole. Eluted proteins were dialyzed against a buffer containing 20 mM Tris (pH 8), 50 mM NaCl, 1 mM BME, and 0.2 mM EDTA. After dialysis, CXCL1/CXCL2 fusion proteins were cleaved using TEV protease (~ 1 mg of enzyme to cleave 50 mg of fusion protein) digestion for 14 hrs at 25 ºC to remove Trx-His fusion tag. Digested proteins were then separated from the Trx tag using S-column pre-equilibrated with buffer containing 20 mM Tris (pH7) and 50 mM NaCl. The non-specific protein and tag impurities were washed using a gradient of NaCl buffers (50 mM and 100 mM). Proteins were eluted in 500 mM NaCl. Further, the digested and undigested fractions of CXCL1/CXCL2 proteins were separated using 2nd Ni-NTA purification. Protein of interest was collected in the flow through and the pure CXCL1/CXCL2 proteins were obtained by passing it through the Superdex-75 size exclusion chromatography (SEC) column. The purity of the proteins was assessed using 15

% SDS-PAGE and the concentration of protein at different purification steps was quantified by performing bicinchoninic acid (BCA) assay.

3.2.3 Size exclusion chromatography (SEC)

SEC was performed on GE health care AKTA prime FPLC system equipped with zinc lamp for absorbance at 215 nm sensitive for amide bond detection and by using HiLoad 16/60 Superdex 75 prep grade column at a flow rate of 1 ml/min. Both CXCL1 and CXCL2 proteins (1 ml each) at two different concentrations (1 mg/ml, and 0.1 mg/ml) were loaded on to the column that was pre-equilibrated with buffer containing 50 mM sodium phosphate (pH 6.0), 100 mM NaCl, and 1 % Glycerol. Further, to assess the oligomerization properties of CXCL1, and CXCL2, other standard reference proteins (pepsin, chymotrypsin, cytochrome C and aprotinin) with known molecular weight were loaded under the same experimental conditions at a concentration of 1 mg/ml.

3.2.4 Glutaraldehyde cross linking assay

Glutaraldehyde cross linking assay [25] was performed by preparing dilutions (1 %, 0.5 %, 0.1 %, 0.05 %) of 25 % glutaraldehyde (sigma) in Milli Q water. For cross linking experiments, 1 mg/ml protein samples of CXCL1 and CXCL2 in 100 mM sodium phosphate buffer incubated with different concentration (0.0005 %, 0.001 %, 0.005 %, 0.01 %) of glutaraldehyde at 25 ºC for 24 hours. Reaction was stopped by adding SDS sample loading buffer into the reaction mixture. All the samples were analyzed using 15 % SDS PAGE. Intensities of monomer, dimer, and tetramer peaks observed for CXCL1 and CXCL2 were calculated using Image J software [26].

3.2.5 Contact map analysis

A comparative contact map depicting Cα contacts in CXCL1 and CXCL2 has been generated using contact map view (CM view) software [27] with distance threshold of 6 Å. 3D coordinates for CXCL1 and CXCL2 were taken using the NMR data as described elsewhere [21,24].

3.2.6 NMR spectroscopy

NMR sample preparation

For NMR studies, ^{15}N- and ^{13}C, ^{15}N- labeled samples of CXCL1 and CXCL2 were exchanged with 50 mM sodium phosphate buffer, pH 6.0 with 0.01 % sodium azide. For three dimensional NMR experiments, ^{13}C, ^{15}N- labeled samples were concentrated to ~ 150 µM for CXCL1 and to ~ 500 µM for CXCL2. To study the homo-oligomerization of CXCL1 and CXCL2, an equal concentration of 150 µM was used and for hetero-dimerization, 60 µM of ^{15}N-CXCL1 and unlabeled CXCL2 were mixed in the ratio of 1:1.

For protein-GAG complex studies, GAGs were either purchased or obtained from research collaborators. Heparin hexasaccharide (HP6, mol. wt. 1800 D) was purchased from Iduron, UK. As per the manufacturer's specification, the main disaccharide unit present in Hp6 is IdoUA,2S–GlcNS,6S (approx 75 %) with some degrees of variations in sulfation patterns. Neocarradodecaose: hexasulfate sodium salt (NC6, mol. wt. 2467 D, catalogue No. N8143) was purchased from sigma. Hyaluronic acid (HA6, mol. wt. 1155 D) and sulfated hyaluronic acid (SHA6, mol.wt. 2570 D) were synthesized by Dr. Sebastian Köhing, university of Berlin. Stocks of 10 mM for all the GAGs were prepared in 50 mM sodium phosphate buffer, pH 6.0.

NMR data acquisition and processing

All the NMR experiments were carried out using a triple channel Bruker 800 MHz/500 MHz spectrometer equipped with a TXI cryoprobe with pulse-shaping and pulse field gradient capabilities. For backbone resonance assignments, standard three dimensional NMR experiments including HNCA, HNCACB, HNCO, and CBCACONH were carried out at 25 ºC.

For CXCL1- 3D NMR experiments, ^{1}H and ^{15}N carrier frequencies were set to 4.7 ppm and at 119.5 ppm respectively. ^{13}C carrier frequencies of 52 ppm for HNCA; 44 ppm for HNCACB, CBCA(CO)NH; and 174 ppm for HNCO were used. For CXCL2, 3D-NMR experiments, ^{1}H and ^{15}N carrier frequencies were set to 4.7 ppm and 120.50 ppm for all the experiments and ^{13}C carrier frequency was set to 53 ppm for HNCA; 42 ppm for HNCACB, CBCA(CO)NH; and 174 ppm for HNCO. Homo-oligomerization of CXCL1 and CXCL2, and their heterodimer formation was monitored by recording ^{15}N-HSQC experiments with 128 scans and 128 complex increments at carrier frequency of 4.7 ppm and 119.5 ppm for ^{1}H and ^{15}N respectively.

For protein:GAG complex studies, ^{1}H-^{15}N HSQC experiments were carried out by concentrating CXCL1, CXCL2 to 60 µM. and CXCL1/2 heterodimer to 30 µM in the apo form and in complex with different GAGs including HP6, HA6, NHA6, and NC6. The final protein-GAG molar concentrations were in the ratio of 1:10. All the NMR experiments were carried out at 298K. All the NMR data was processed in Bruker Topspin 3.5, and subsequent assignment analysis was carried out in CARA [28].

3.2.7 Generation of homo-/hetero-oligomeric structures

Structures for CXCL2 (PDB ID: 1MI2, 3N52) were available in RCSB protein data bank. Structural model for CXCL1 and CXCL3 were generated as described in **Section 2.2.7**, **Chapter 2**. Homo-/Hetero-dimers were generated through the monomeric counterparts of CXCL1/CXCL2/CXCL3 structures as described in **Section 2.2.7, Chapter 2**.

3.3 Results

3.3.1 Cloning and expression of CXCL1 and CXCL2

To unravel the homo and hetero-oligomerization features of GRO chemokines (CXCL1, CXCL2) and to study the effect of GAG binding on chemokine oligomerization, CXCL1 and CXCL2 chemokines were cloned, expressed, and purified. The amplified CXCL1 and CXCL2 genes from pMD18-T cloning vector have been successfully ligated using T4 ligase between the Kpn1 and Nco1 restriction sites of pET32 expression vector containing thioredoxin (Trx) protein as a fusion moiety to enhance the solubilization of protein of interest and hexa histidine affinity tag that aids in facile purification of the fusion protein. A TEV protease cleaving sequence (ENLYFQG) was inserted between the fusion protein and the gene of interest in order to facilitate the Trx and His removal of tags **(Fig. 3.1)**. The amplified products of CXCL1 and CXCL2 genes have been shown in the **Fig. 3.2 A**. The successful cloning of the CXCL1 and CXCL2 gene into pET32 has been confirmed by DNA sequencing result **(Fig. 3.2 B)**. To check the expression of CXCL1 and CXCL2, *E. coli* [BL21 (DE3)] cells were transformed with recombinant CXCL1/CXCL2 plasmids and induced with IPTG. The control and the overexpressed cells were lysed and the protein bands were analyzed using 12 % SDSPAGE. The PAGE gel evidenced for an over expressed band of 25 kD at the expected size of CXCL1/CXCL2 fusion protein **(Fig. 3.2 C)**.

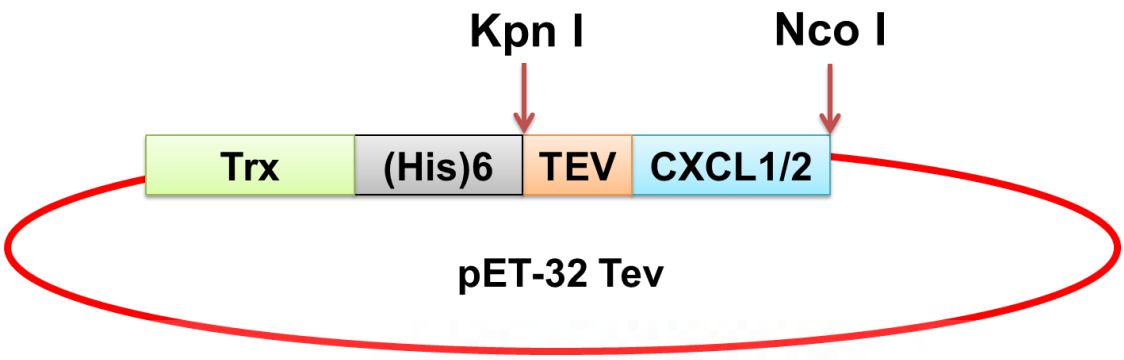
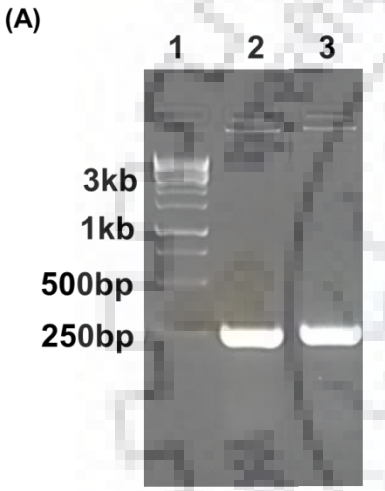
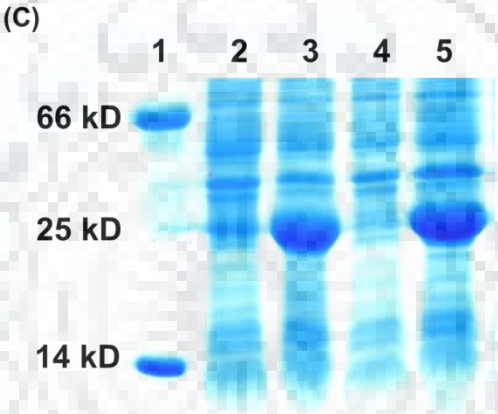


Figure 3.1: *Schematic of pET-32 vector showing the positioning of the fusion tag, restriction site, TEV protease cleavage site, and CXCL1/CXCL2 gene in their respective order.*





(B)

CXCL1

```
Ggt aac gaa aac ctg tat ttt cag ggg ggt gct ccg att gct aac gaa ctg cgt tgt caa
 G   T   E   N   L   Y   F   Q   G   G   A   P   I   A   N   E   L   R   C   Q
tgc ctg caa acc atg gcc ggt atc cat ctg Aaa aac atc caa agc ctg aag gtc ctg ccg
 C   L   Q   T   M   A   G   I   H   L   K   N   I   Q   S   L   K   V   L   P
agc ggt ccg cat tgc acc cag acg gaa gtc att gcg acc ctg aaa aac ggt cgt gaa gcg
 S   G   P   H   C   T   Q   T   E   V   I   A   T   L   K   N   G   R   E   A
tgt ctg gat ccg gaa gcc ccg ctg gtg cag aaa atc Gtt caa aaa atg ctg aag ggc gtt
 C   L   D   P   E   A   P   L   V   Q   K   I   V   Q   K   M   L   K   G   V
ccg aag taa cca ttg
 P   K   -   P   W
```

CXCL2

```
Ggt aac gaa aac ctg tat ttt cag ggg gct gtc gtt gcg agt gaa ctg cgt tgc caa tgt
 G   T   E   N   L   Y   F   Q   G   A   V   V   A   S   E   L   R   C   Q   C
ctg aaa acc ctg ccg cgt gtt gac ttt aaa aat atc caa tct ctg agc gtg acc ccg ccg
 L   K   T   L   P   R   V   D   F   K   N   I   Q   S   L   S   V   T   P   P
ggt ccg cat tgc gca cag acc gaa gtg att gcc acg ctg aaa ggc ggt caa aaa gtt tgc
 G   P   H   C   A   Q   T   E   V   I   A   T   L   K   G   G   Q   K   V   C
ctg gac ccg gaa gcc ccg ctg gtc caa aaa atc atc caa aaa atc ctg aac aaa ggc aaa
 L   D   P   E   A   P   L   V   Q   K   I   I   Q   K   I   L   N   K   G   K
gca aac taa cca ttg
 A   N   -   P   W
```

Figure 3.2: (A) *1.5 % agarose gel showing the 1 kB DNA ladder in lane 1, and an amplified CXCL1 and CXCL2 genes (255 base pairs) in lane 2 and 3 respectively,* **(B)** *Sequencing result (255 base pairs) confirming the presence of CXCL1 and CXCL2 nucleotide sequence (black) and protein sequence (red) along with TEV cleavage site (green) and restriction sites Kpn1 at N-terminus and Nco1 at C-terminus (blue). The amino acid sequence for the encoded gene is presented with single letter code,* **(C)** *12 % SDS-PAGE analysis of expression of CXCL1 and CXCL2, Lanes 1, 2 and 4 represent marker, un-induced samples of CXCL1 and CXCL2, Lanes 3 and 5 represent induced samples of CXCL1 and CXCL2.*

3.3.2 Purification of CXCL1 and CXCL2 chemokines

CXCL1/CXCL2 proteins were purified by the combination of chromatography techniques (affinity, ion exchange and SEC). The purity at every step of the purification process was assessed by SDS-PAGE, and the yields were determined from BCA assay **(Fig. 3.3,** and **Table 3.1).** The purification protocol involved four steps: (a) In the first step, affinity chromatography was performed using Ni-NTA to bind the fusion protein using its $(His)_6$ Tag. Most of the non-specific impurities were washed with a gradient of imidazole and the fusion protein (~ 25 kD) was eluted in various fractions **(Fig. 3.3 A1/B1)**. (b) In the second step, the fractions containing the fusion protein were pooled, twice dialyzed to remove the imidazole, and the resultant protein was subjected to TEV enzyme cleavage. TEV digestion resulted in three fractions; a small amount of uncut fused protein (~ 25 kD), Trx-His tag (~ 17 kD) and the CXCL1/CXCL2 protein (~ 8 kD) **(Fig. 3.3 A2/B2)**. (c) The CXCL1/CXCL2 protein (pI – 8.5) was then separated from fusion tag (pI – 5.1) using cation exchanger (SP-sepharose). The elution profiles of CXCL1/CXCL2 using 500 mM NaCl contain mainly CXCL1/CXCL2 protein and some amount of the fused protein **(Fig. 3.3 A3/B3)**. (d) CXCL1/CXCL2 was then separated from the undigested fusion protein using a 2^{nd} Ni-NTA step. As CXCL1/CXCL2 does not have a tag on its own, it will not bind to Ni-NTA and has been collected in flow through **(Fig. 3.3 A4/B4)**. An overall yield of ~ 12 mg of CXCL1 and 20 mg of CXCL2 pure recombinant protein has been obtained by exploiting the Trx fusion protein expression and purification protocol in a bacterial expression system. Further, the purity of proteins was enhanced by gel filtration chromatography.

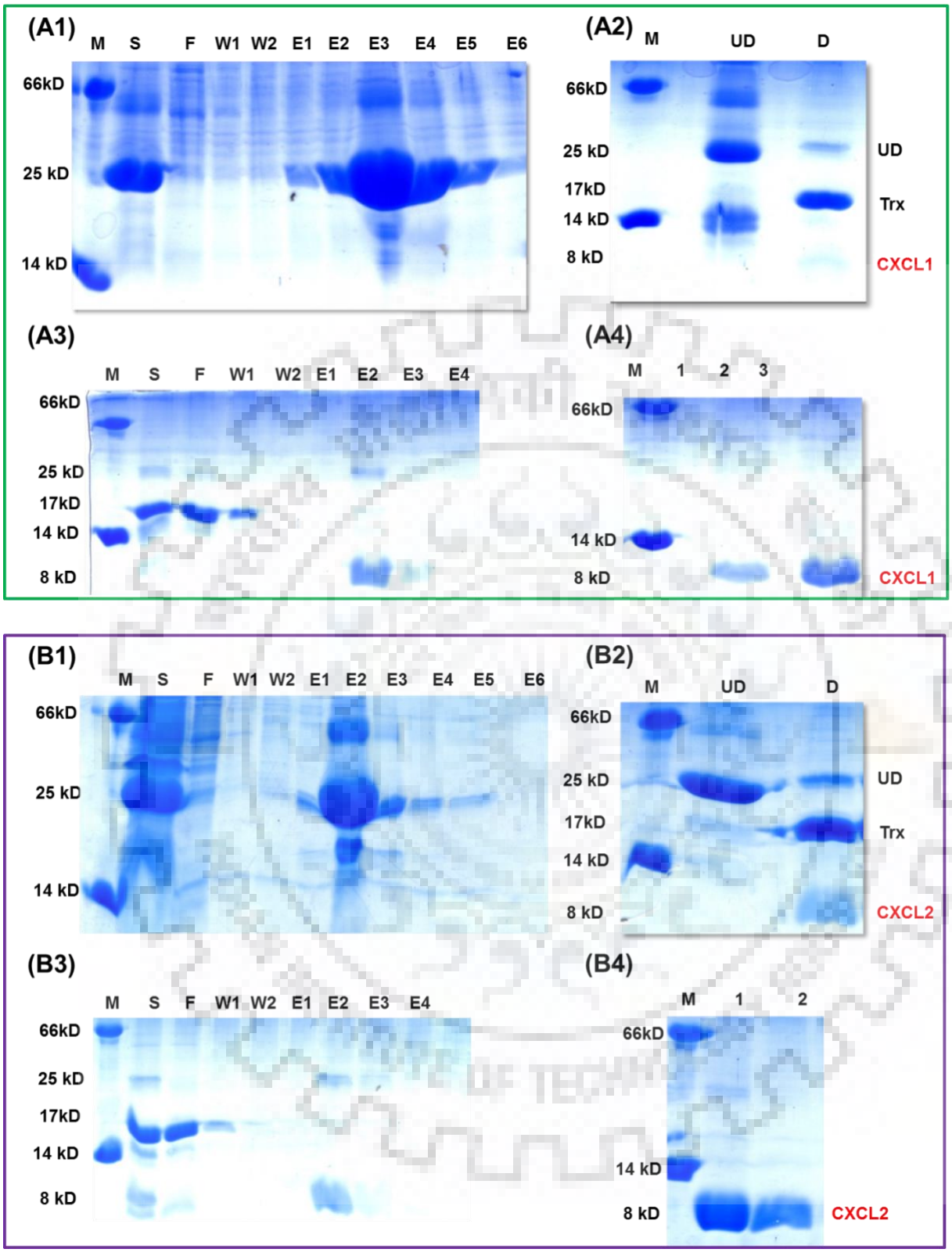


Figure 3.3: *15 % SDSPAGE gel analysis of protein migration at different purification steps for CXCL1 in upper (green), and CXCL2 in lower (purple) panel. In both panels,* **(A1/B1)** *1st Ni NTA purification step: Lane S-Bacterial cell lysate, Lane F-flow through fraction, Lanes W1 and W2-washes with 10 mM and 30 mM imidazole, Lanes E1, E2, E3, E4, E5, E6, E7-eluted fractions using 400 mM imidazole;* **(A2/B2)** *Tev Digestion step: Lane UD–undigested sample*

of 25 kD, Lane D-digested sample with three bands of 25 kD, 17 kD and 8 kD represents the left over undigested sample, thioredoxin tag and digested CXCL1/CXCL2 protein respectively; **(A3/B3)** *Tag removal step using cation exchange chromatography: Lane S; undigested sample as supernatant, Lane F-flow through fraction; Lanes W1 and W2-washes with 50 mM NaCl and 100 mM NaCl respectively, Lanes E1, E2, E3, E4-eluted fractions using 500 mM NaCl.* **(A4/B4)** *Reverse Ni-NTA chromatography step: Lanes 1, 2, and 3- flow through fractions containing 8 kD CXCL1/CXCL2 protein. In all the gels, Lane M- marker containing two proteins lysozyme (14 kD) and BSA (66 kD).*

Table 3.1: Summary of the amount of CXCL1 and CXCL2 proteins obtained from 1 L of LB culture after each step of purification.

Purification step	CXCL1 (mg)	CXCL2 (mg)
1st Ni-NTA purification	100 ± 10	150 ± 10
Dialysis	70 ± 5	95 ± 5
Cation exchange chromatography	20 ± 2	30 ±2
Reverse Ni NTA purification	15 ± 2	25 ± 2
Gel filtration chromatography	12 ±2	20 ± 1

3.3.3 Oligomerization potencies of CXCL1 and CXCL2

The oligomeric state of CXCL1 and CXCL2 was assessed by comparative analysis of their elution profiles from size exclusion chromatography. The chromatograms of both CXCL1 and CXCL2 at higher concentration (1 mg/ml) indicated that the proteins are eluted at slightly different volumes (CXCL1 - 73 ml, and CXCL2 - 71.5 ml) **(Fig. 3.4 A)**. The elution profile of CXCL1 and CXCL2 were compared with standard proteins of different sizes including pepsin (36 kD), chymotrypsin (25.6 kD), cytochrome C (12 kD) and aprotinin (6 kD). The elution volumes evidenced that both CXCL1 and CXCL2 were eluted at a retention time that nearly corresponds to apparent molecular weight of ~ 16 kD, thus confirming the presence of dimeric species at the chosen concentration. CXCL1 and CXCL2 at lower concentrations (0.1 mg/ml) were eluted at different elution volumes as compared to their elutions at higher concentration. The difference in the elution volumes of CXCL1 and CXCL2 at higher and lower concentration is around 3 ml for CXCL1 and is 1 ml for CXCL2 **(Fig. 3.4 B-C)**. Such a differential shift of the peak position between CXCL1 and CXCL2 indicates that the majority of CXCL1 at lower concentration is present as monomer with only a small fraction remained as intact dimer. Whereas, in CXCL2, the majority is existing in the dimeric state and only small amount of it is dissociated in monomeric form. The presence of a single peak for CXCL1/CXCL2 at low concentration and a peak shift instead of appearance of new peak at

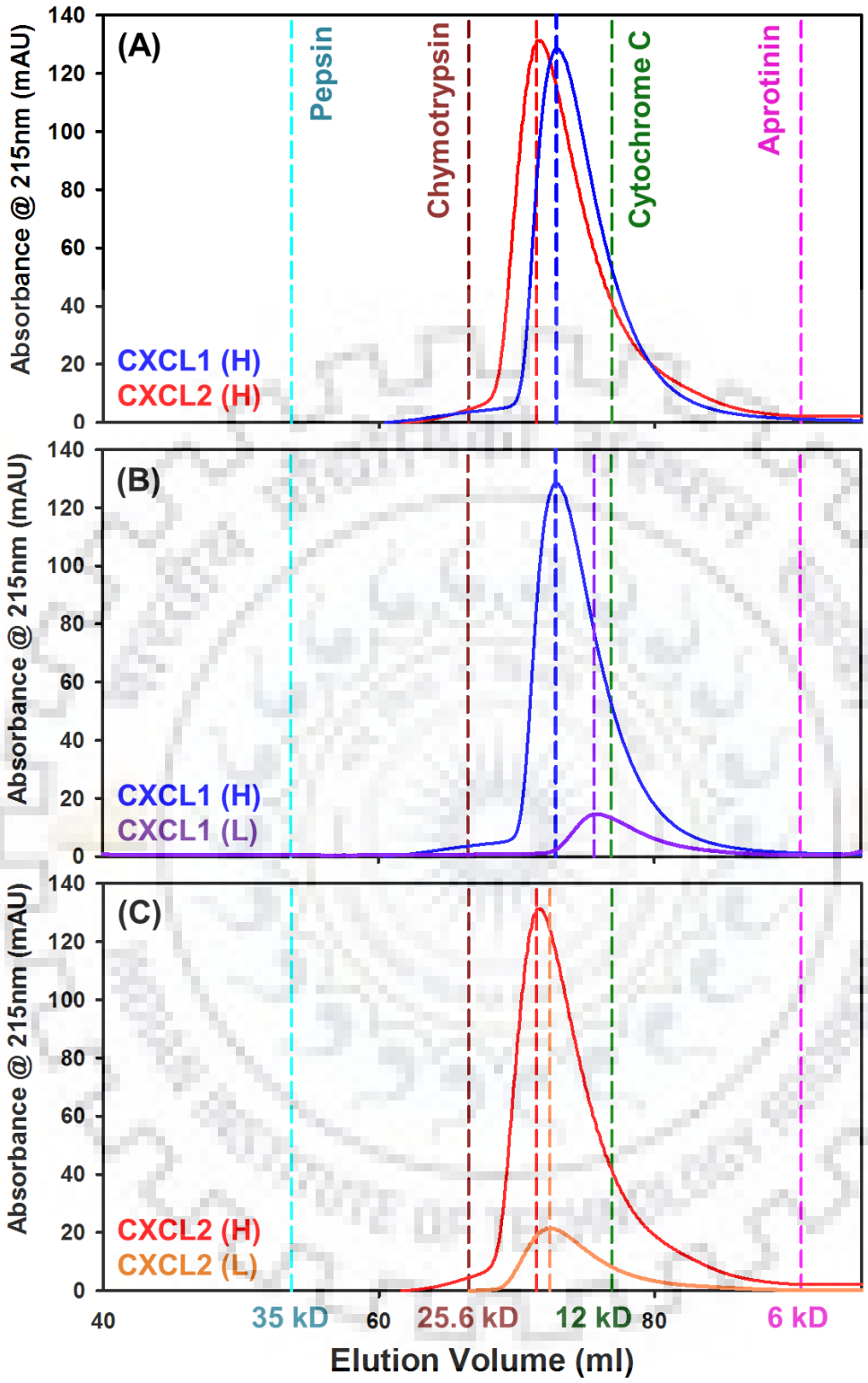


Figure 3.4: *Size exclusion chromatography elution profiles of* **(A)** *CXCL1 (blue) and CXCL2 (red)* at higher *concentrations* **(B)** *CXCL1 at higher (blue) and lower (purple) concentrations* **(C)** *CXCL2 at higher (red) and lower (yellow) concentrations. Standard marker proteins of different sizes are shown as straight lines at their elution maxima. H and L denote higher (1 mg/ml) and lower (0.1 mg/ml) concentration of proteins.*

monomeric position suggests that the both the species are in fast exchange in the SEC experimental time scale. These results provide the first line of evidence for the strong dimeric nature of CXCL2 as compared to CXCL1.

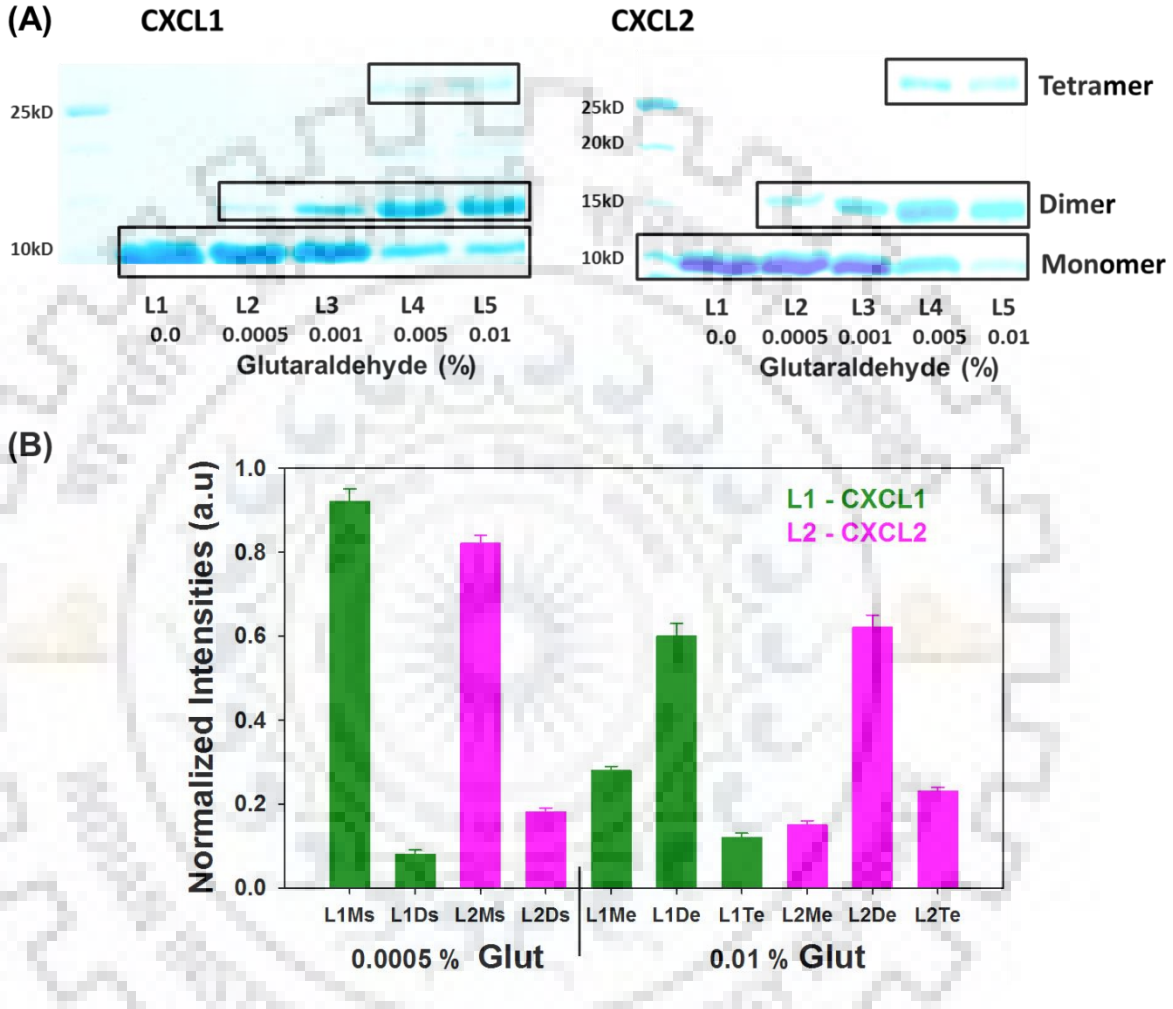


Figure 3.5: (A) *15 % SDS-PAGE analysis of glutaraldehyde cross linking of CXCL1 and CXCL2;* **(B)** *Normalized intensities of CXCL1, and CXCL2 oligomeric forms generated by 0.005 %and 0.01 % glutaraldehyde; M, D, T denotes monomer, dimer, and tetramer oligomeric states; 's' and 'e' represents starting and ending concentrations of glutaraldehyde.*

Further, their oligomerization affinities under the influence of chemical cross linker were analyzed. Glutaraldehyde cross linking assay was used to examine the oligomerization potentials of CXCL1 and CXCL2 under same experimental conditions including protein concentrations. As shown in **Fig. 3.5 A**, under the influence of the crosslinking agent both proteins forms higher order oligomers. At lower concentration of glutaraldehyde (0.0005 % and 0.001 %) the dimeric species are formed, and a further increase of the cross-linker

concentration (0.005 % and 0.01 %) resulted in higher order tetrameric species. In order to delineate their oligomerization efficacies, the intensity profiles of the various oligomeric states were quantified **(Fig. 3.5 B)**. The intensity analysis suggested that CXCL2 has a higher affinity to form the oligomers as compared to CXCL1. At 0.0005 % of glutaraldehyde an excess of monomeric species of CXCL1 is left over and at higher concentration (0.01 % glutaraldehyde), although overall dimeric content of both proteins are similar, the ratio of the left over monomer to tetramer varies suggesting that CXCL1 has a weak oligomerization tendency as compared to its paralogue CXCL2 **(Fig. 3.5 B)**.

The oligomerization profiles observed in the crosslinking experiment depends on two factors (a) the association constant, which indeed is a ratio of the forward and backward kinetic rate constants and; (b) nature and availability of the residues that are essential for crosslinking. In order to confirm that the observed differential homo oligomerization behavior is not due to the influence of covalent cross-linker, but an inherent and unique feature of their association rates, ^{1}H-^{15}N HSQC experiments of CXCL1 and CXCL2 at same concentration (~ 150 µM) were performed **(Fig. 3.6)**. Clearly, CXCL1 displayed two set of peaks corresponding to its dimeric (major) and monomeric (minor) populations in contrast to single set of dimeric resonances in CXCL2 spectrum, establishing their intrinsic differential homo dimerization capabilities. Such a differential oligomerization can be attributed to the dynamic nature of the C-terminal helix in CXCL1 as compared to CXCL2 in the dimeric conformation [21,24].

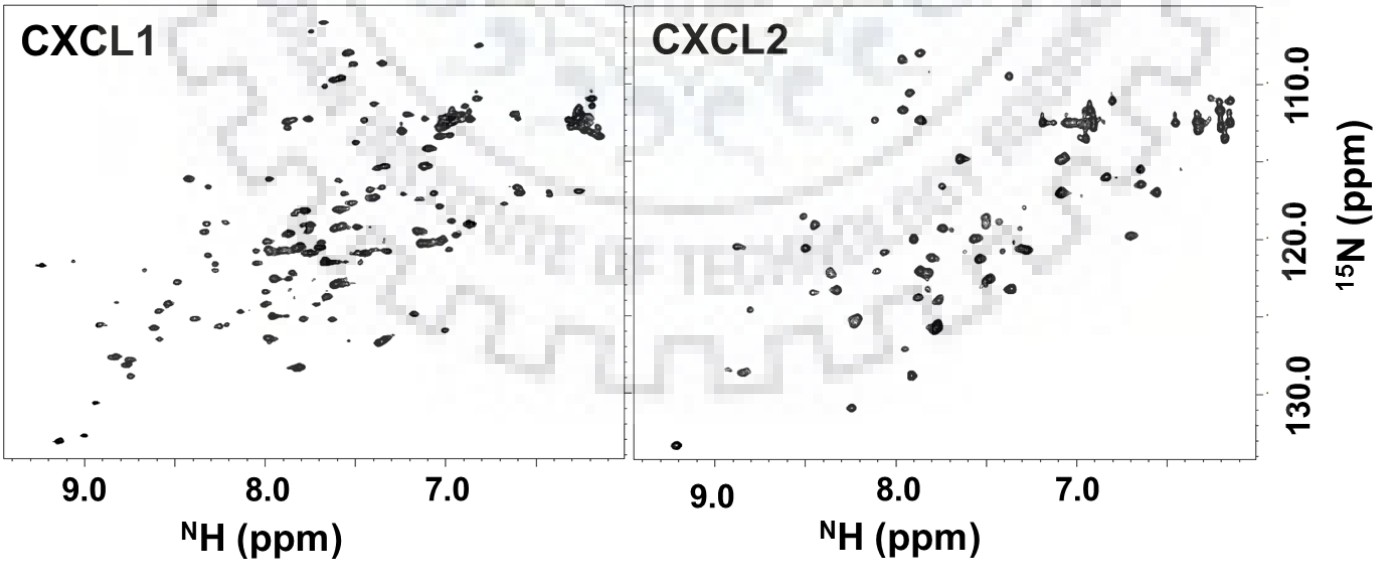


Figure 3.6: *^{1}H-^{15}N HSQC spectra of CXCL1 and CXCL2 (~ 150 µM) at 25 ^{o}C.*

3.3.4 Resonance assignments of CXCL1 and CXCL2

In order to characterize the residues from the monomeric and dimeric species in CXCL1 and the residues participating in the dimer interface of CXCL1 and CXCL2, the NH cross peaks in the HSQC spectra of CXCL1 and CXCL2 were assigned using conventional triple resonance experiments as described in **Section 3.2.4**. In CXCL1, some of the residues in the non-overlapping regions of monomer and dimer were assigned. The HNCACB strips for one of such stretch including residues (T15-G18) from both monomer and dimer species of CXCL1 have been shown in **Fig. 3.7**.

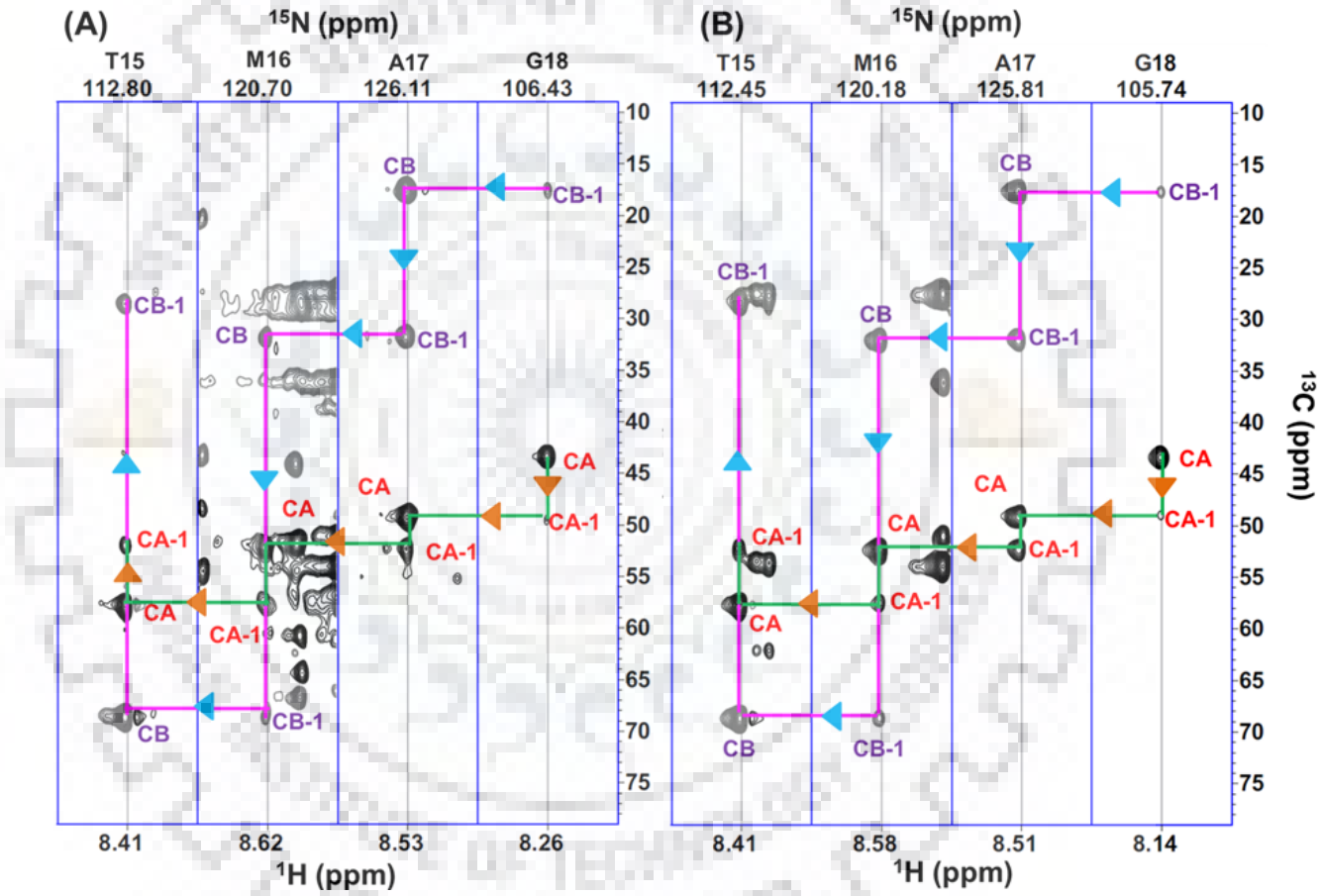


Figure 3.7: *HNCACB strip plot for a stretch of four residues (T15-M16-A17-G18) showing the connectivities of Cα resonances with pink lines and blue arrow heads and of Cβ resonances with green lines and yellow arrow heads for monomer* **(A)** *and dimer* **(B)** *species of CXCL1.*

Backbone NH assignments for the assigned residues for both the monomer and dimer species have been marked in **Fig. 3.8**.

For CXCL2 dimer, Out of total 73 amino acids, 6 are proline that are not observed in the spectra and hence, rest all the 67 amino acids that are observed in the spectra have been assigned. HNCACB strips used for the assignment have been shown for the stretch of four consecutive amino acids (A4-R8) **(Fig. 3.9)**. Backbone NH assignments for the assigned residues of CXCL2 have been shown in **Fig. 3.10**.

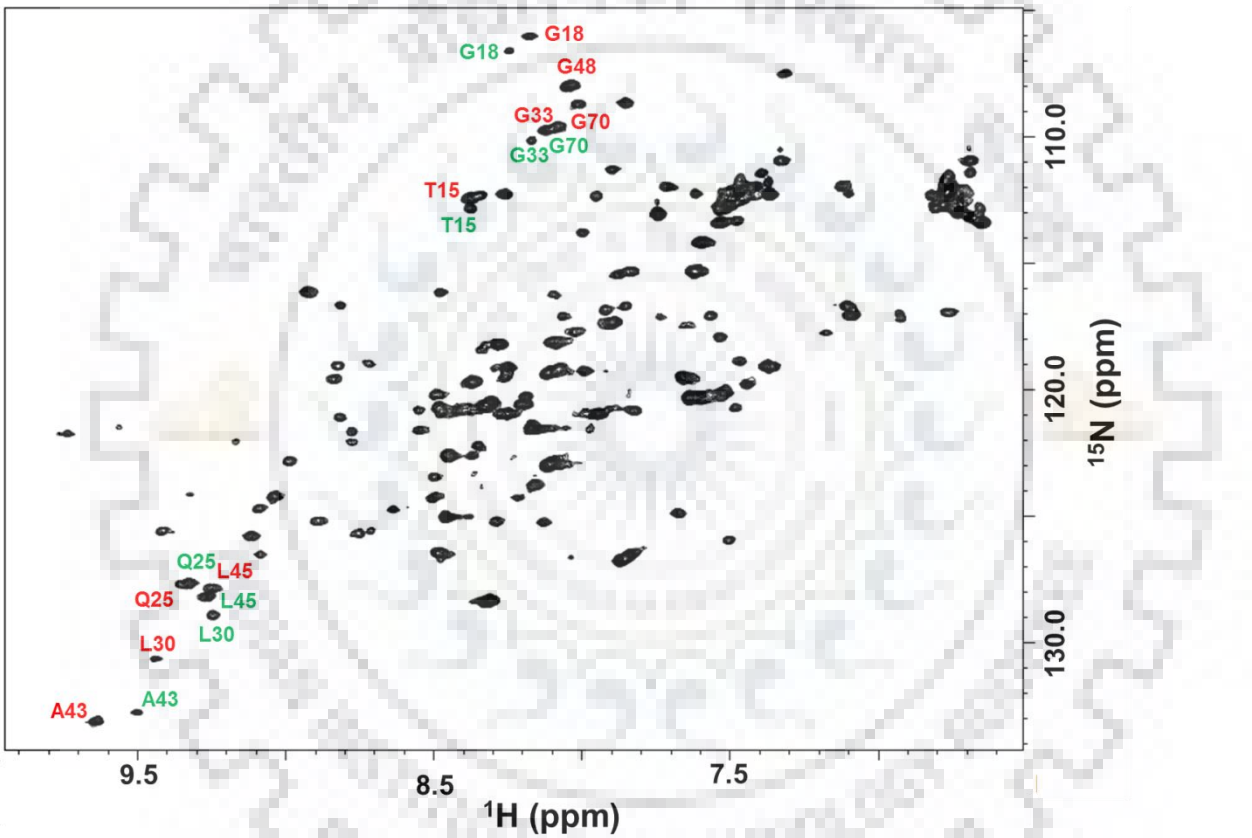


Figure 3.8: *^{1}H-^{15}N HSQC spectrum showing the residue specific resonance assignments for CXCL1 monomeric species (green) and dimeric species (red). The NH cross peaks are labeled with residue symbol and number.*

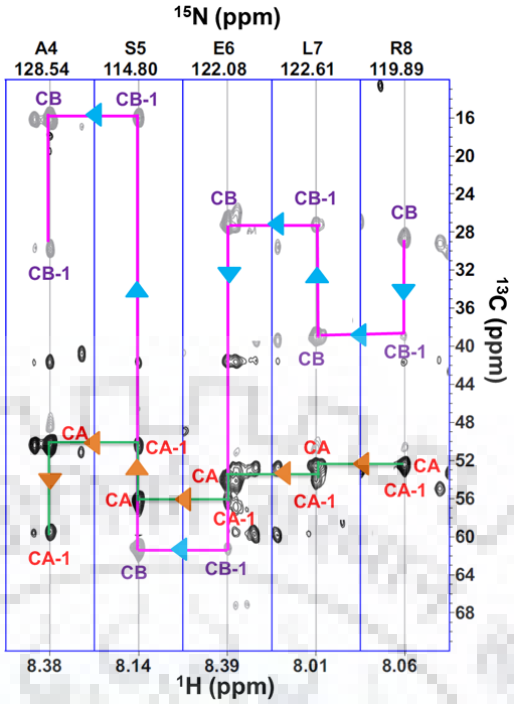


Figure 3.9: *HNCACB strip plot for a stretch of four residues (A4-S5-E6-L7-R8) showing the connectivities of C_α resonances with pink lines and blue arrow heads and of C_β resonances with green lines and yellow arrow heads for CXCL2.*

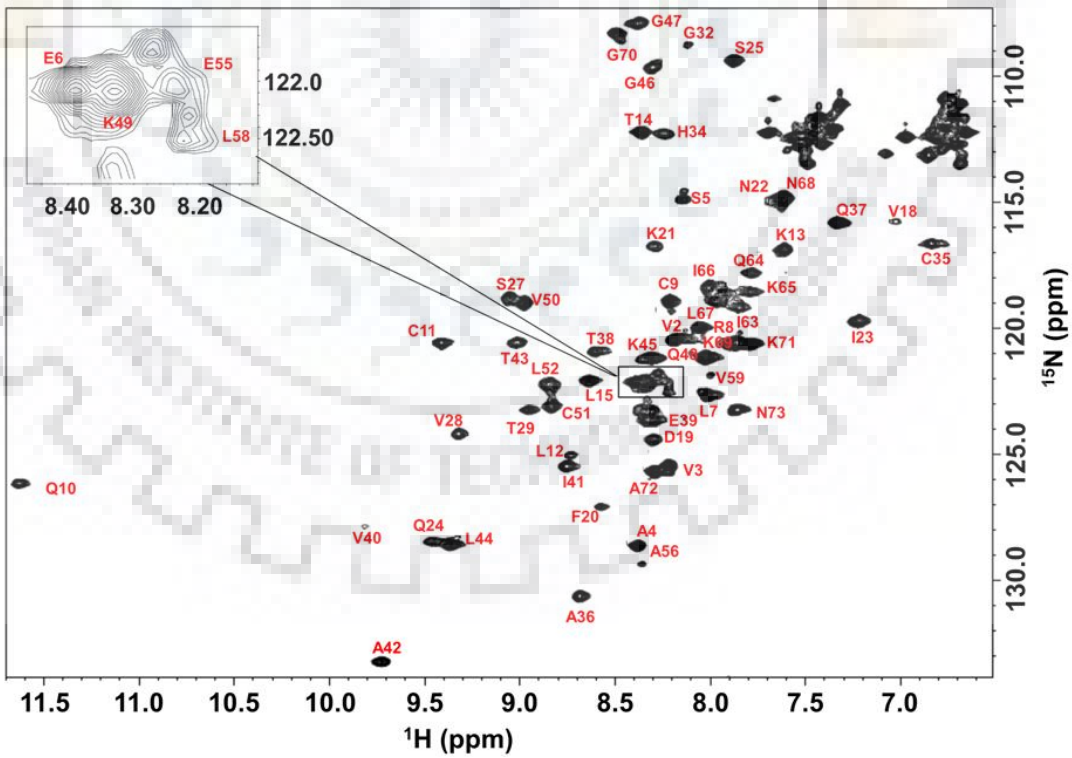


Figure 3.10: *^1H-^{15}N HSQC spectrum showing the residue specific resonance assignments for CXCL2. The NH cross peaks are labeled with residue symbol and number. Central peaks have been shown in zoomed box enclosed with in the spectra.*

3.3.5 In silico structural characterization of CXCL1 and CXCL2

Recent structural studies of the CXCL1 dimer evidenced the similarity of structural features of CXCL1/CXCL2 in their monomeric and dimeric forms [21,22]. Upon confirming the differential homo dimerization and similar structural features, the plausibility of heterodimer formation within these two entities sharing high sequence and structural similarity was examined. First, comparative analysis of C_α contacts for CXCL1 and CXCL2 was carried out by constructing their contact maps **(Fig. 3.11 A)**. An overlap of 82 % was observed in the contact maps of CXCL1 and CXCL2 and the remaining 18 % of the differences in the contacts are due to the differences in their local amino acid sequences. Further, sequence analysis suggested that, at the β_1-β_1' strand of dimeric interface **(Fig. 3.11 B)**, only a single residue is different (L30 in CXCL1 and T29 in CXCL2). Thus, all the in-silico analysis is pointing towards a favorable heterodimer formation.

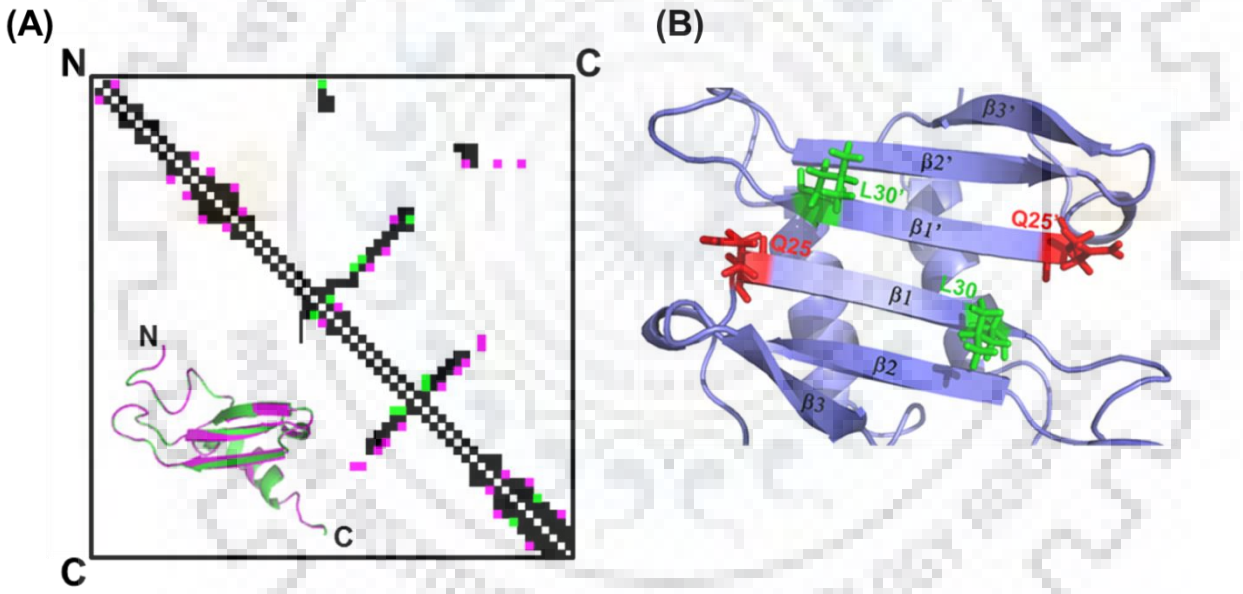


Figure 3.11: (A) *Overlay of contact maps of CXCL1 (green) and CXCL2 (pink) where N and C represents N-terminal and C-terminal of the protein sequences, the structural overlay of the monomeric subunits are also shown;* **(B)** *Structure of CXCL1 dimer highlighting the residues at the dimer interface β-strand termini.*

3.3.6 Heterodimerization of CXCL1 and CXCL2

In order to establish the formation of CXCL1/2 heterodimers experimentally, an equimolar concentrations of both proteins were mixed and monitored the ^{1}H-^{15}N resonances of CXCL1 **(Fig. 3.12)**. The addition of unlabeled CXCL2 resulted in new set of dimer-interface peaks **(Fig. 3.12, 3.13 A)** along with the existing two sets of resonances (CXCL1

monomer/homo-dimer), accompanied by a significant attenuation in the peak intensities. The intensities of the dimer interface residues Q25 and L30 of CXCL1 were quantified **(Fig. 3.13 B)** before and after the addition of CXCL2, in order to calculate the populations of the three species (monomer, homo and hetero dimers). Analysis suggested the presence of 30-35 % of monomeric species and 60-65 % of dimer species (in terms of monomeric population) in the apo spectra, which is in good agreement with the reported K_d (~ 30 μM) for CXCL1 monomer-dimer equilibrium [21]. The homo monomer-dimer equilibrium is now distributed in the ratio of ~ 25 % monomer, ~ 35 % homodimer, and 40 % heterodimer for both the resonances Q25 and L30 in the complex spectra **(Fig. 3.13 B)**, thus directly demonstrating the potential formation of CXCL1-CXCL2 heterodimers.

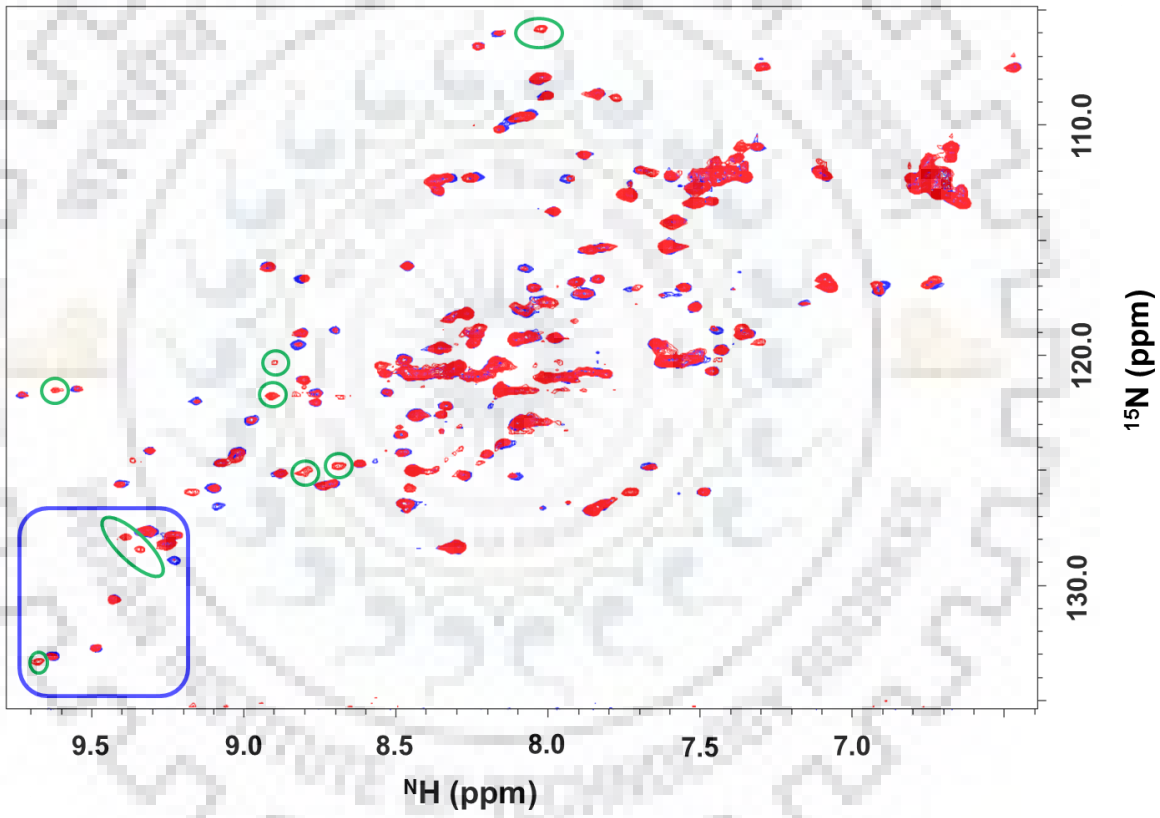


Figure 3.12: *Overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-CXCL1 (blue) and ^{15}N-CXCL1+^{14}N-CXCL2 (red) in the ratio of 1:1 showing the interaction of CXCL1 with CXCL2. Some of the well resolved resonances appeared in the complex spectra upon formation of the heterodimer are encircled in green. The spectral region comprising of monomeric, dimeric, and heterodimeric peaks was encircled in black.*

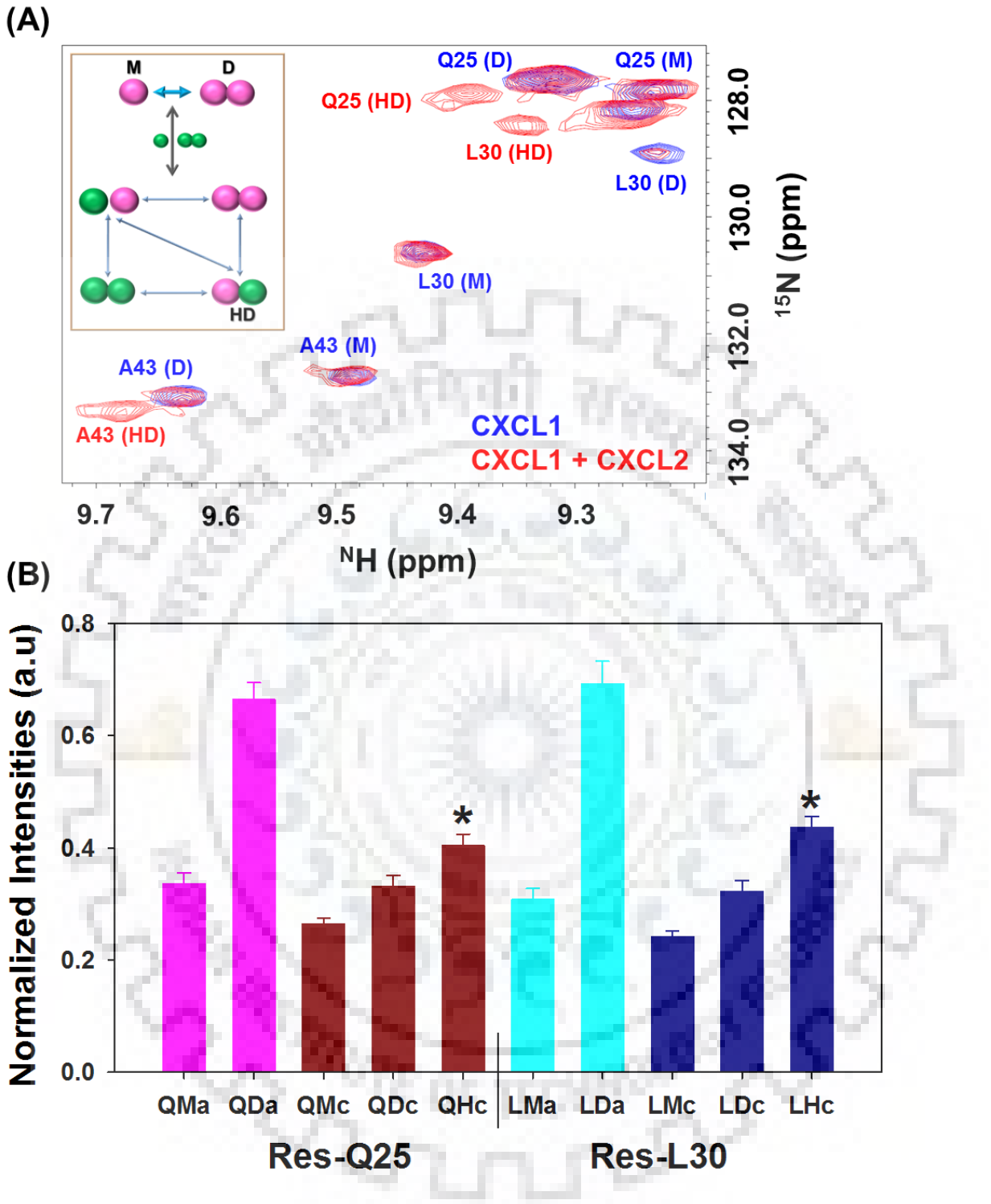


Figure 3.13: (A) *Overlay of ^{1}H-^{15}N HSQC spectra [spectral region encircled in black (Fig. 3.12)] of ^{15}N-CXCL1 (blue) and ^{15}N-CXCL1+^{14}N-CXCL2 (red) in the ratio of 1:1 showing the molecular interaction of CXCL1 with CXCL2. The notations in the inset are M-monomer, D-Dimer, and HD-hetero dimer;* **(B)** *Normalized intensities of the NH resonances of Q25 and L30; M, D, H represents monomer, dimer, hetero-dimer respectively; 'a' and 'c' represents alone (CXCL1) and complex (CXCL1+CXCL2) respectively; * denotes the intensities of hetero-dimer peaks specific to the CXCL1/2 complex spectra.*

3.3.7 Effect of GAG binding on homo- and hetero-dimerization of CXCL1 and CXCL2

To probe the effect of GAG binding on the homo- and hetero-oligomerization behavior of CXCL1 and CXCL2, the ^{1}H-^{15}N HSQC experiments for CXCL1, CXCL2, and their heterodimer both in the absence and the presence of GAGs were performed. Lower concentration of proteins with CXCL1/CXCL2 (60 µM), and CXCL1/2 heterodimer (30 µM), was used in order to ensure the presence of monomeric species and to clearly demonstrate the effect of GAG binding on the oligomerization of the proteins. Four different GAGs namely heparin hexasaccharide (HP6), hyaluronan hexasaccharide (HA6), synthetic sulfated hyaluronan hexasaccharide (known as SHA6), and Neocarradodecaose: hexasulfate sodium salt (NC6) to assess the GAG induced oligomerization of CXCL1 and CXCL2 **(Fig. 3.14)**. Hexasaccharide GAGs were used to assure the efficient locking of the GAGs between the C-terminal helices of both the monomeric units in the dimer. Theoretically calculated distance between the two helices indicated that the minimum length of six saccharide units of GAGs is required to bind perpendicularly to the helices for a tight dimer formation [21]. The purpose of using four different GAGs/GAG mimetics is as follows. HP6 is a natural GAG that is present in tissues and animal cells, which is commonly used to assess the GAG binding properties of chemokines/other GAG binding immunoregulatory proteins. HA6 is also a natural GAG, and lacks sulfation. SHA6 is a synthetic HA6/GAG mimetic with uniform sulfation. SHA6 can mimic HP6 and can be a potential therapeutic target. NC6 is a natural marine GAG with limited sulfation, i.e., only one sulfate group per disaccharide unit. NC6 can also be probed for its potential therapeutic to regulate natural chemokine-GAG interactions.

Fig. 3.15 shows the superposition of spectra of ^{1}H-^{15}N HSQC spectra of CXCL1 in the absence (blue) and in the presence of different types of GAGs (red). **Fig. 3.15 A** clearly shows two sets of CXCL1 resonances in the absence of GAG, indicating the presence of both the monomeric and the dimeric species. After the addition of HP6, only one set of resonances corresponding to the dimeric species of CXCL1 was observed. This establishes that binding of HP6 induced dimerization in CXCL1. No such dimerization effect was observed in CXCL1 upon the addition of HA6, implying that there is no interaction between CXCL1 and HA6 **(Fig. 3.15 b)**. In contrast to the result of HA6, dimerization effect was observed in CXCL1 upon the addition of SHA6 **(Fig. 3.15 C)**. Dimerization of CXCL1 in the presence of sulfated GAGs

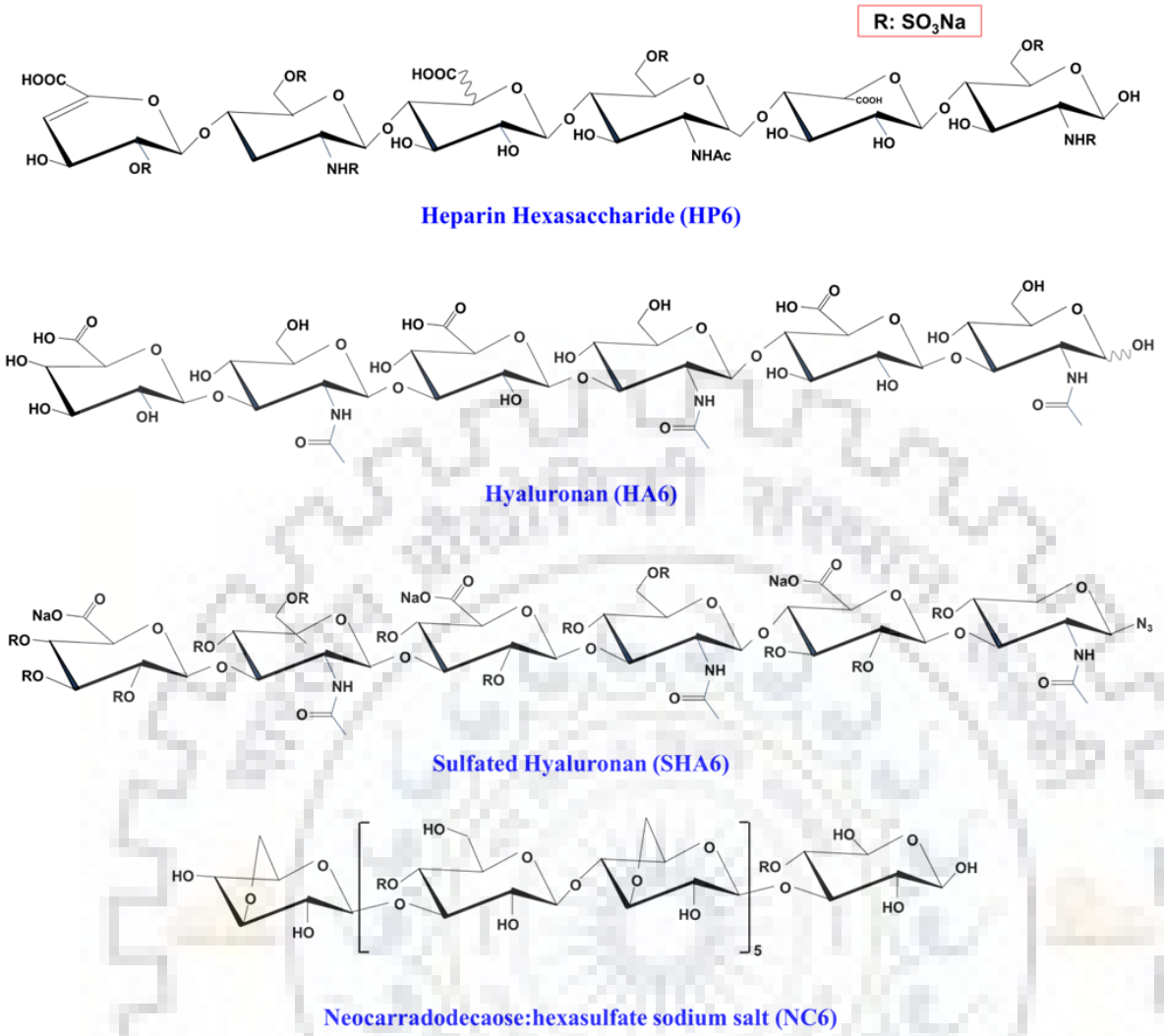


Figure 3.14: *Structure for the different GAGs employed to elucidate the effect of GAGs on homo- and hetero-oligomerization of chemokines.*

(HP6 and SHA6), but not with HA6 clearly exemplifies the importance of sulfate groups in chemokine-GAG interactions. To further get insights into the role of sulfate groups on GAGs for chemokine-GAG interactions and GAG induced dimerization, NC6, a limited sulfated GAG was used. No dimerization of CXCL1 was observed upon the addition of NC6, despite of presence of one sulfate group per disaccharide in NC6 **(Fig. 3.14, 3.15 D)**. This result indicates that the presence of sulfate groups alone is not sufficient for GAGs to interact with chemokines, but the position and extent of sulfation pattern in GAGs also play a significant role in mediating chemokine-GAG interactions/GAG induced oligomerization.

Similar ^{1}H-^{15}N HSQC experiments for CXCL2 with GAGs were also carried out to decipher the fact that these GAG/GAG mimetic interactions are exclusively specific for

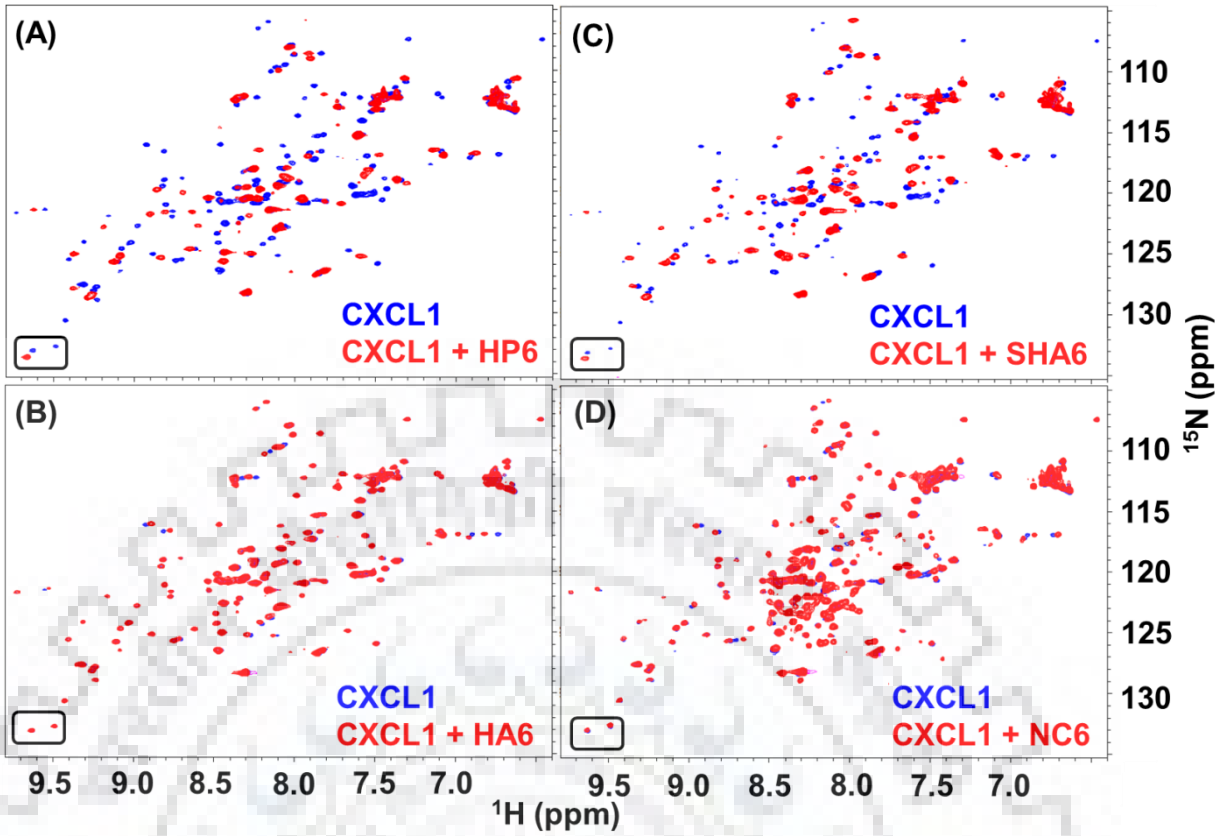


Figure 3.15: *Overlay of ^{1}H-^{15}N HSQC spectra of CXCL1 in the apo form (blue) and in the presence of different GAGs (red) including* **(A)** *heparin hexasaccharide (HP6),* **(B)** *hyaluronan hexasaccharide (HA6),* **(C)** *sulfated hyaluronan hexasaccharide (SH6) and* **(D)** *Neocarradodecaose: hexasulfate sodium salt (NC6). The spectral region comprising of monomeric, and dimeric peak of A43 are encircled in black.*

CXCL1 or they hold true for other NAC/GRO chemokine homodimers **(Fig. 3.16)**. Similar GAG induced oligomerization effect as that of CXCL1 was observed with CXCL2 indicating that these interactions hold good for the GRO chemokine subfamily of CXC chemokine homodimers.

Further to understand the effect of these GAGs/GAG mimetics on hetero-oligomerization of CXCL1 and CXCL2, ^{1}H-^{15}N HSQC experiments were also performed for CXCL1/2 heterodimer. **Fig. 3.17** shows the superposition of CXCL1/2 heterodimer spectra in the absence (blue) and in the presence of GAGs (red). Spectra for CXCL1/2 heterodimer in the absence of GAGs clearly show three set of resonances as described above. After addition of HP6 or SHA6, GAG induced dimerization effect was observed as the three sets of resonances collapsed into one/two sets comprising of both the homo and heterodimeric species **(Fig. 3.17 A and C)**. No dimerization effect was observed upon HA6 or NC6 addition to CXCL1/2 heterodimer as similar to CXCL1 and CXCL2 **(Fig. 3.17 B and D)**. In order to demonstrate the

observed changes in CXCL1, CXCL2, and CXCL1/2 heterodimer after the addition of different GAGs/GAG mimetics, **Fig. 3.18** shows the zoomed view of resonance of A43, from both monomer and dimer species in CXCL1 **(panel A)**, and CXCL1/2 heterodimer spectra **(panel C)** and resonance of A42 in CXCL2 **(panel B)**. The third peak corresponding to heterodimer peak of A43 is also observed in CXCL1/2 heterodimer spectra **(Fig. 3.18, panel C)**. All the results clearly established the variable dimerization effects induced by different GAGs. These results demonstrate that GAG induced oligomerization holds true for both homo and hetero oligomerization of NACs. The summary of GAG induced dimerization behavior of CXCL1, CXCL2, and CXCL1/2 with different GAGs is provided in **Table 3.2.**

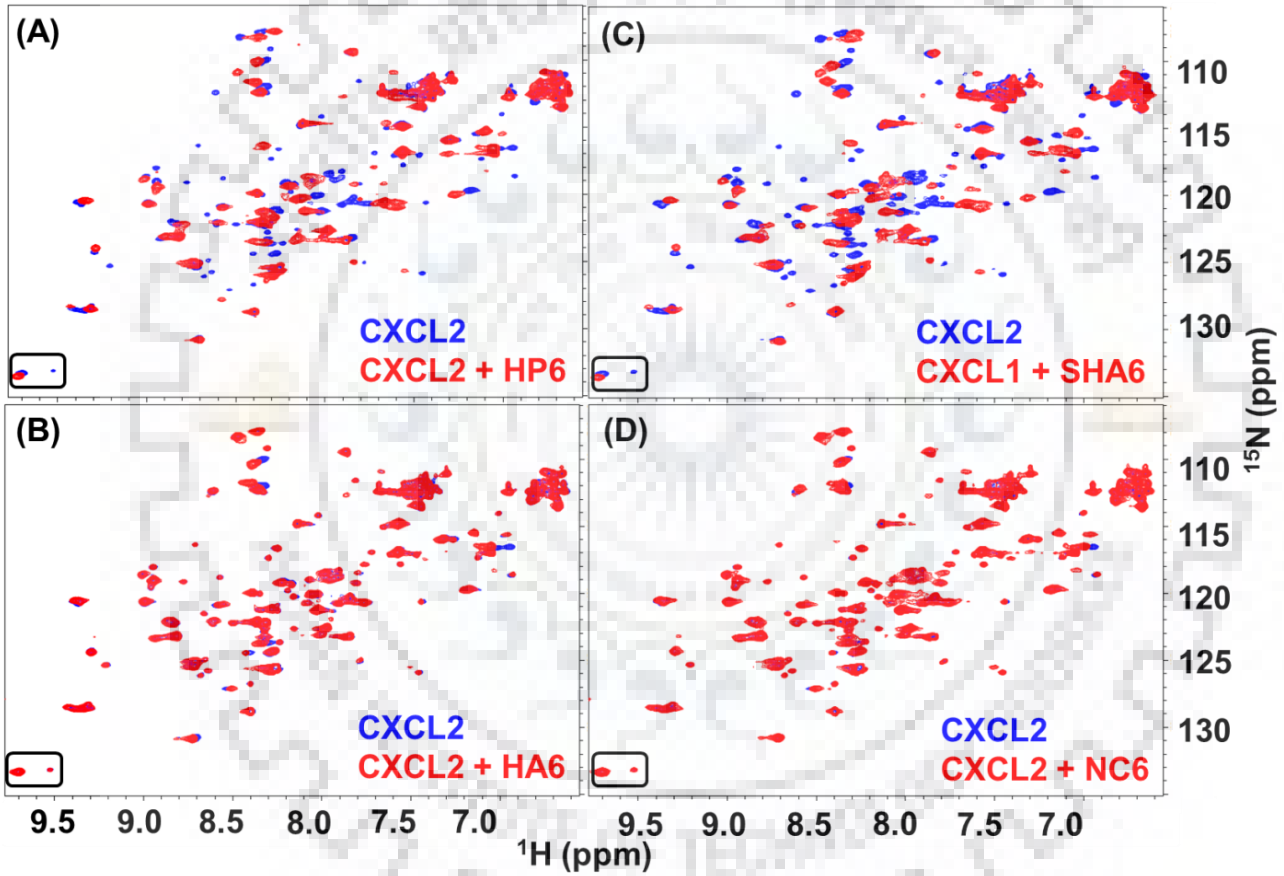


Figure 3.16: *Overlay of ^{1}H-^{15}N HSQC spectra of CXCL2 in the apo form (blue) and in the presence of different GAGs (red) including* **(A)** *heparin hexasaccharide (HP6),* **(B)** *hyaluronan hexasaccharide (HA6),* **(C)** *sulfated hyaluronan hexasaccharide (SH6) and* **(D)** *Neocarradodecaose: hexasulfate sodium salt (NC6). The spectral region comprising of monomeric, and dimeric peak of A42 are encircled in black.*

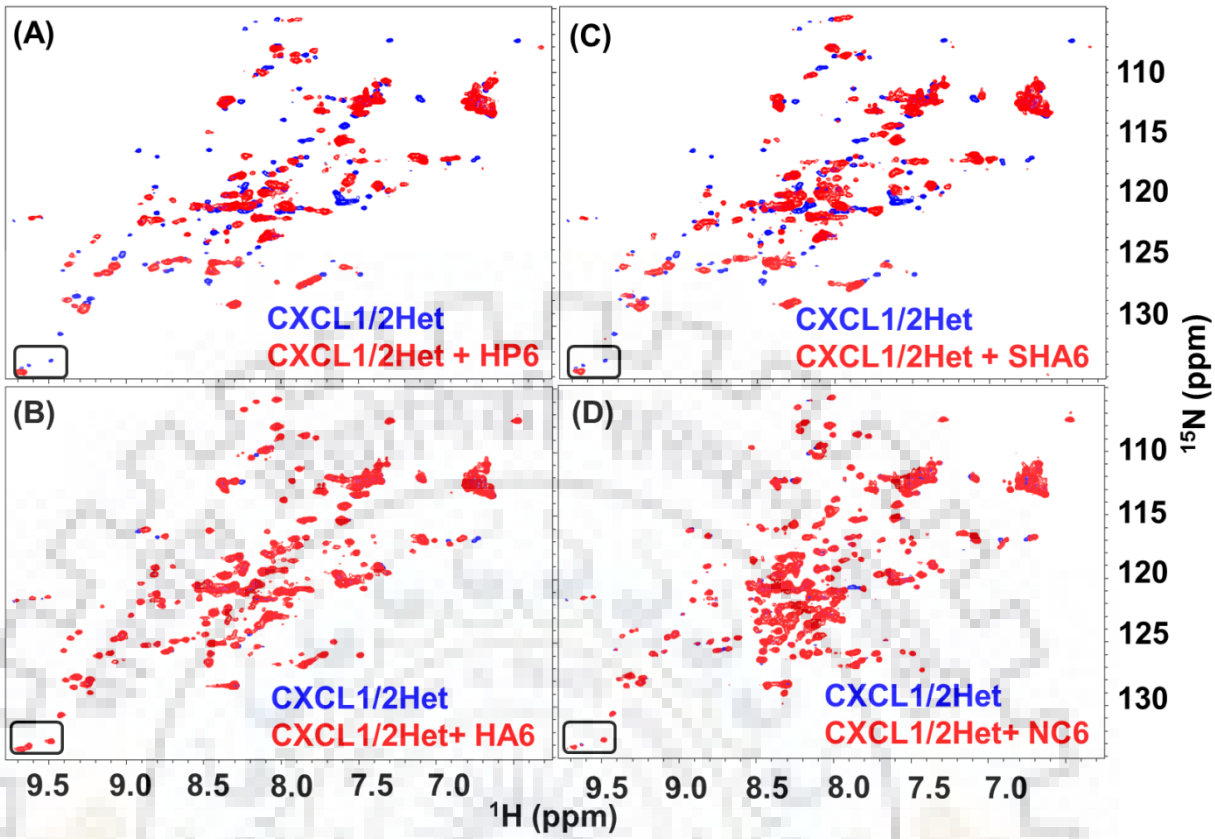


Figure 3.17: *Overlay of ^{1}H-^{15}N HSQC spectra of CXCL1/CXCL2 heterodimer in the apo form (blue) and in the presence of different GAGs (red) including* **(A)** *heparin hexasaccharide (HP6),* **(B)** *hyaluronan hexasaccharide (HA6),* **(C)** *sulfated hyaluronan hexasaccharide (SH6) and* **(D)** *Neocarradodecaose: hexasulfate sodium salt (NC6). The spectral region comprising of monomeric, dimeric, and heterodimeric peak of A43 are encircled in black.*

Table 3.2: *Summary of the effect of different GAGs on GAG induced dimerization in CXCL1, CXCL2 and the CXCL1/2 heterodimer. The tick mark (✓) indicates the GAG induced dimerization and the cross mark (✗) implies no binding interaction.*

Type of GAG	CXCL1	CXCL2	CXCL1/CXCL2
Heparin Hexasaccharide (HP6)	✓	✓	✓
Hyaluronan (HA6)	✗	✗	✗
Sulfated Hyaluronan (SHA6)	✓	✓	✓
Neocarradodecaose: hexasulfate sodium salt (NC6)	✗	✗	✗

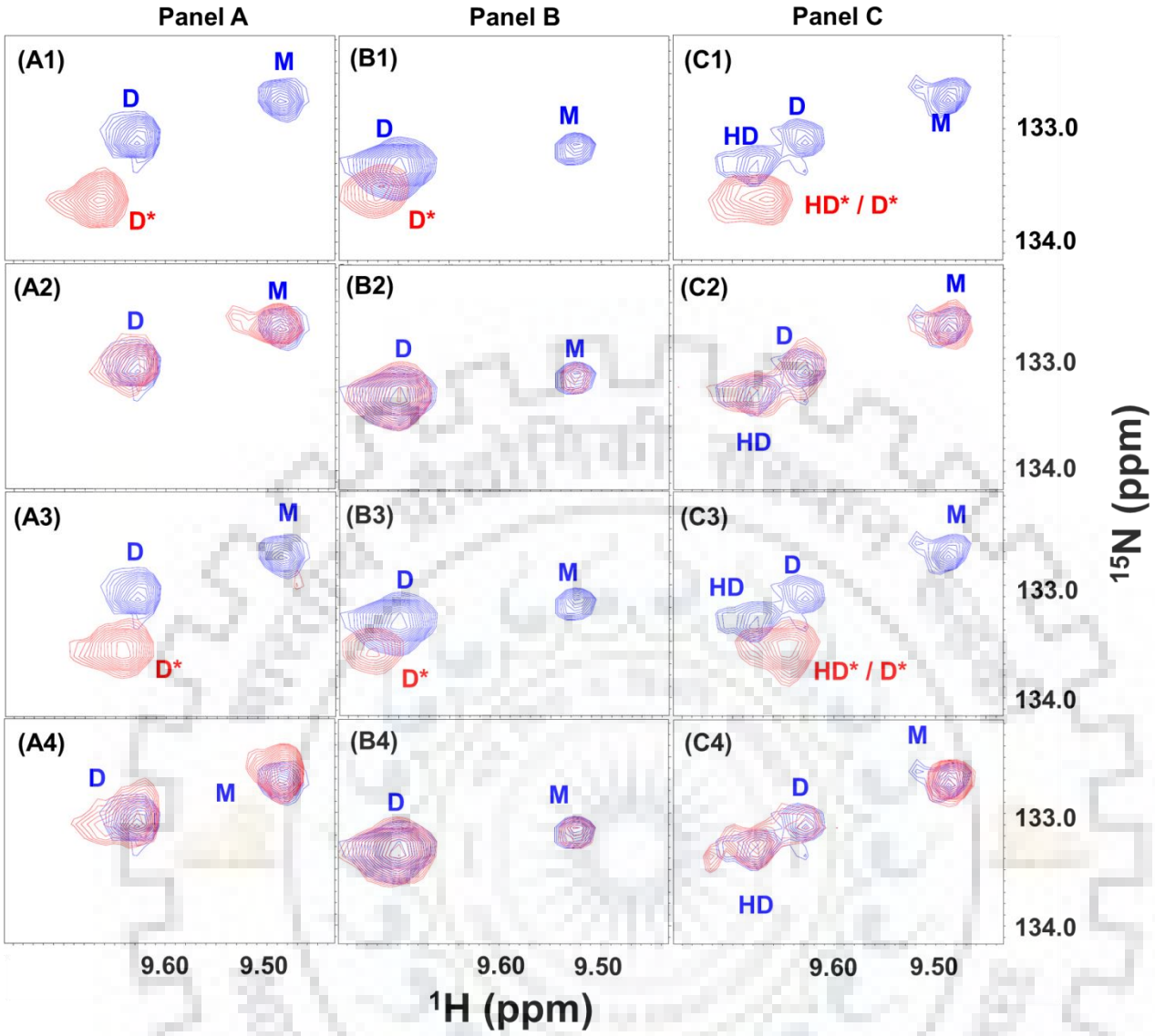


Figure 3.18: *Zoomed view of spectral region of CXCL1* **(panel A)**, *CXCL2* **(panel B)**, *and CXCL1/CXCL2 heterodimer* **(panel C)** *showing the monomeric (M), dimeric (D) and heterodimeric (HD) resonances of A43 (in CXCL1 and CXCL1/2 heterodimer) and of A42 (in CXCL2) in the absence (blue) and presence of different GAGs (red). The numbers (1, 2, 3, 4) in each of the panel indicates the spectra in the presence of HP6, HA6, SHA6, and NC6 respectively. Star notation in the inset indicates the resonance arose as a result of GAG induced homo/ heterodimerization.*

3.4 Discussion

As discussed in **Chapter 2**, GRO chemokines (CXCL1, CXCL2 and CXCL3) are formed as a resultant of the gene duplication events **(Fig. 3.19 A)** with a high level of sequence and structural similarity with varied functional potencies in all the mammalian species [21,22,24,29-32]. The varied functional behaviors can be attributed to; (i) the differential receptor and GAG binding surfaces, where the key residues are altered during evolution thus

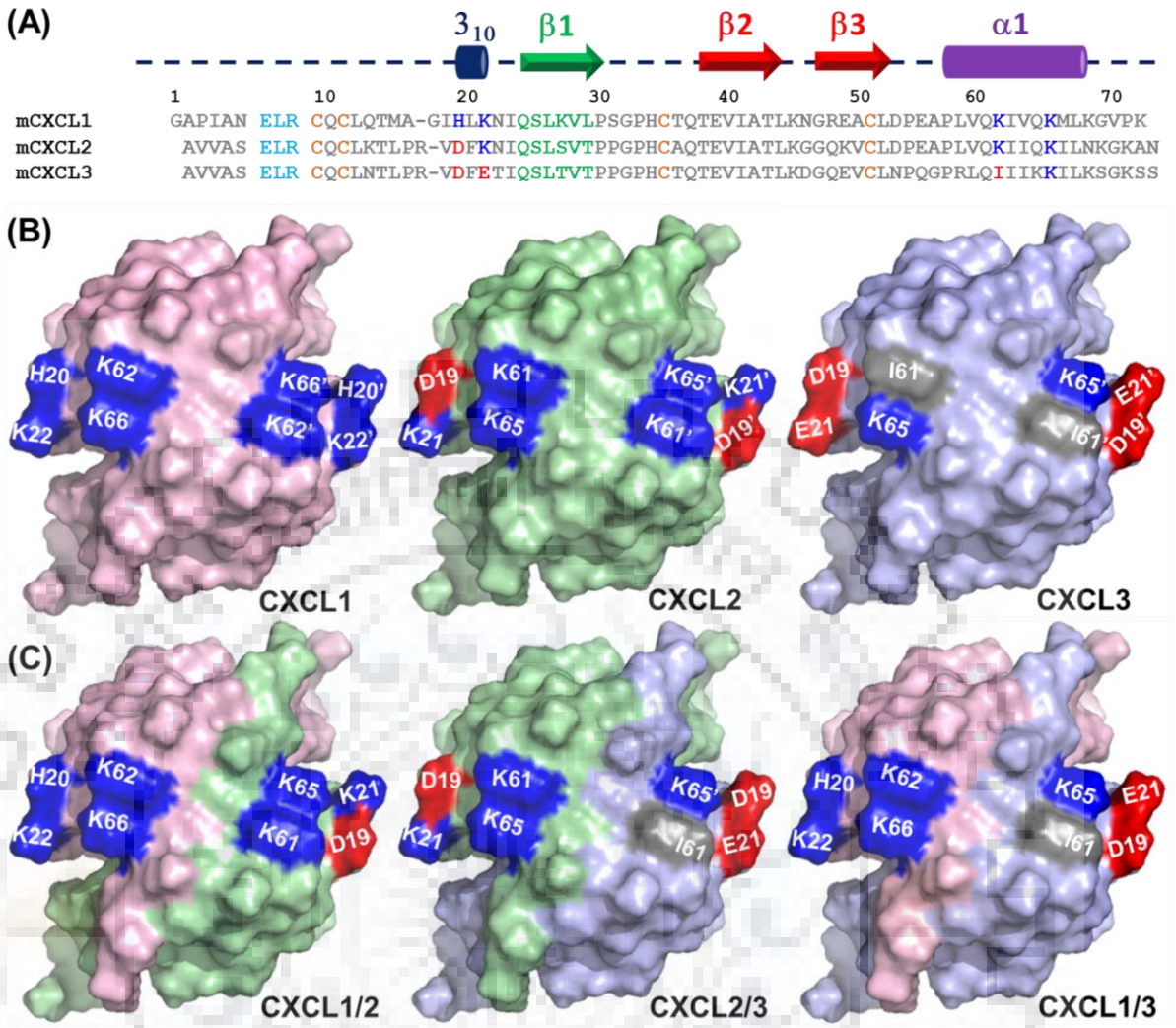


Figure 3.19: (A) *Amino acid sequences of murine GRO chemokines (mCXCL1, mCXCL2 and mCXCL3) in which conserved residues are highlighted (ELR motif - Cyan, cysteine involved in disulfide bridges – Yellow), positive charge residues H20, K22, K62, K66 crucial for GAG binding (blue) and their alterations (red). The secondary structural elements are shown with arrows (β1 essential for dimerization - green, β2 and β3- red) and cylinders (310helix - blue, α-helix-purple); Surface representation of homodimers* **(B)** *and heterodimers* **(C)** *of CXCL1 (light pink), CXCL2 (pale green) and CXCL3 (slate blue). The monomeric fold of the CXCL3 was generated through homology modeling, and the homo/hetero dimeric structures of all the three chemokines were obtained by performing the symmetry operations in Pymol molecular graphics system as described in supplementary methods. The essential residues for GAG binding on the helical surface and their alterations are annotated and highlighted. Positive charge residues (blue), negative charge residues (red) and hydrophobic residues (grey).*

creating the same structure with differential binding surfaces. In the case of CXCL1, the GAG binding residues are recognized to be H20, K22, K62, and K66 respectively [21]. Alteration of these residues in the homodimers will significantly influence the GAG binding and hence the leukocyte recruitment process by the CXCL1/CXCL2/CXCL3 proteins **(Fig. 3.19 A-B)**. (ii)

The variability in the oligomerization equilibrium also significantly contribute to the differential activity/recruitment profile; as the net accumulation of the individual chemokine oligomeric variants potentially regulates the steepness and sustainability of chemotactic gradients [1]. (iii) Further, the recruitment profile is also altered by the generation of hetero-dimeric/oligomeric species in which both the oligomerization propensities and the receptor/GAG binding efficacies attenuate. Such heterodimeric species can generate more contrasting binding surfaces on each monomeric counterpart **(Fig. 3.19 C)**, thus contributing to the specificity and differential binding characteristics.

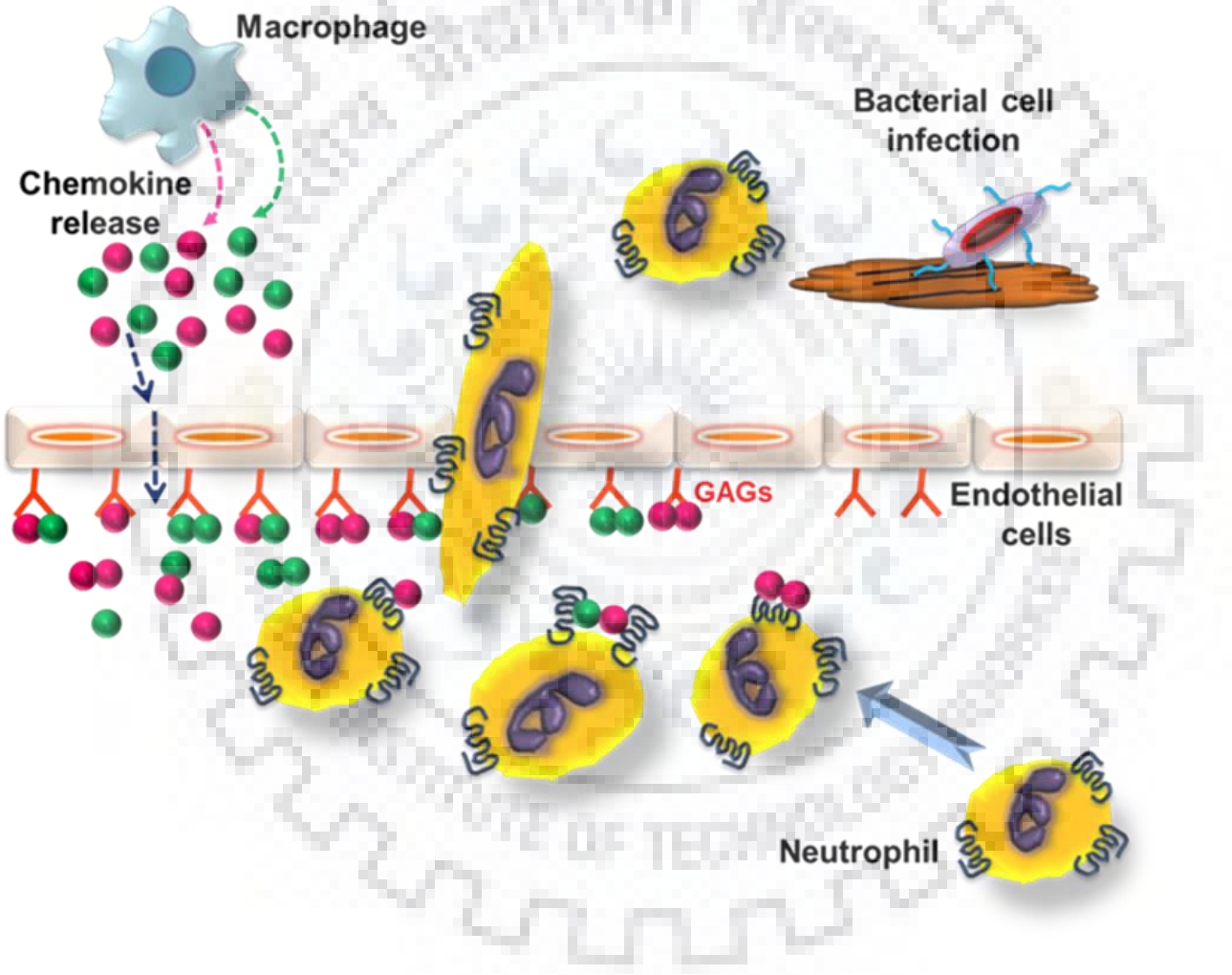


Figure 3.20: *Schematic showing the different oligomeric species (monomers, homodimers and heterodimers) of CXCL1/CXCL2 (pink/green spheres) chemokines driving neutrophil recruitment process at the site of infection, through their interactions with cell surface glycosaminoglycans (GAGs) and G-protein coupled receptors (GPCRs) on the neutrophils.*

These molecular regulatory mechanisms can potentially add another layer of regulatory mechanism in governing the influx of leukocyte migration during infection and injury **(Fig. 3.20)**. Moreover, such hetero-species are less susceptible to protease degradation compared to

their homo-oligomers as the site of cleavage and degradation kinetics varies in both monomeric counterparts. Indeed studies on hetero-oligomers between the CXC chemokine PF4/CXCL4 and the CC chemokine RANTES/CCL5 depicted the enhancement of leukocyte arrest on endothelial cells as compared to CCL5 alone [9]. On a similar note, heterodimers of CXCL4 and CXCL8 had direct implications with an enhancement of CXCL4 anti-proliferative effect on endothelial cells and CXCL8 mediated migration of HCXCR2 expressing neutrophils [11].

However, till now the effect of GAG binding particularly on hetero-oligomerization of NAC chemokines has not been delineated. This study demonstrated that as long as the electrostatic surfaces are closely related and contains the essential GAG binding residues, the GAGs do promote hetero-oligomerization as like homo-oligomerization, although the affinities will be dictated by the total number of GAG interacting residues on a hetero dimeric surface. A quantitative and comparative knowledge of chemokine-GAG binding affinities for both hetero and homo dimeric species using specific GAGs is imperative to substantiate the observed oligomerization based structural characteristics of the heterodimeric chemokines.

3.5 Conclusions

This chapter deciphered the homo/hetero-oligomerization potencies of CXCL1 and CXCL2, and the effect of different GAGs/GAG mimetics on their oligomerization. The oligomerization experiments established that the CXCL1 and CXCL2 form homo and hetero oligomers under in-vitro conditions with different efficacies. GAG binding studies revealed that HP6 and SHA6 induced both homo/hetero-dimerization in CXCL1 and CXCL2 in contrast to HA6 and NC6. These studies evidenced that GAG binding promotes both homo and hetero-dimerization of CXCL1 and CXCL2 chemokines, and the sulfation pattern of GAGs play an important role in defining their interaction with chemokines and henceforth their oligomerization.

3.6 References

1. Salanga CL, Handel TM. Chemokine oligomerization and interactions with receptors and glycosaminoglycans: the role of structural dynamics in function. Exp. Cell Res. 317(5), 590-601 (2011).

2. Wang X, Sharp JS, Handel TM, Prestegard JH. Chemokine oligomerization in cell signaling and migration. Prog. Mol. Biol. Transl. Sci. 117, 531-578 (2013).

3. Zhang X, Chen L, Bancroft DP, Lai CK, Maione TE. Crystal structure of recombinant human platelet factor 4. Biochemistry. 33(27), 8361-8366 (1994).

4. Wang X, Watson C, Sharp JS, Handel TM, Prestegard JH. Oligomeric structure of the chemokine CCL5/RANTES from NMR, MS, and SAXS data. Structure. 19(8), 1138-1148 (2011).

5. Swaminathan GJ, Holloway DE, Colvin RA, Campanella GK, Papageorgiou AC, Luster AD, Acharya KR. Crystal structures of oligomeric forms of the IP-10/CXCL10 chemokine. Structure. 11(5), 521-532 (2003).

6. Murphy JW, Yuan H, Kong Y, Xiong Y, Lolis EJ. Heterologous quaternary structure of CXCL12 and its relationship to the CC chemokine family. Proteins. 78(5), 1331-1337 (2010).

7. Malkowski MG, Wu JY, Lazar JB, Johnson PH, Edwards BF. The crystal structure of recombinant human neutrophil-activating peptide-2 (M6L) at 1.9-A resolution. J Biol. Chem. 270(13), 7077-7087 (1995).

8. Crown SE, Yu Y, Sweeney MD, Leary JA, Handel TM. Heterodimerization of CCR2 chemokines and regulation by glycosaminoglycan binding. J Biol. Chem. 281(35), 25438-25446 (2006).

9. von HP, Koenen RR, Sack M, Mause SF, Adriaens W, Proudfoot AE, Hackeng TM, Weber C. Heterophilic interactions of platelet factor 4 and RANTES promote monocyte arrest on endothelium. Blood. 105(3), 924-930 (2005).

10. Venetz D, Ponzoni M, Schiraldi M, Ferreri AJ, Bertoni F, Doglioni C, Uguccioni M. Perivascular expression of CXCL9 and CXCL12 in primary central nervous system lymphoma: T-cell infiltration and positioning of malignant B cells. Int. J Cancer. 127(10), 2300-2312 (2010).

11. Nesmelova IV, Sham Y, Dudek AZ, van Eijk LI, Wu G, Slungaard A, Mortari F, Griffioen AW, Mayo KH. Platelet factor 4 and interleukin-8 CXC chemokine heterodimer formation modulates function at the quaternary structural level. J Biol. Chem. 280(6), 4948-4958 (2005).

12. Nesmelova IV, Sham Y, Gao J, Mayo KH. CXC and CC chemokines form mixed heterodimers: association free energies from molecular dynamics simulations and experimental correlations. J Biol. Chem. 283(35), 24155-24166 (2008).

13. Guan E, Wang J, Norcross MA. Identification of human macrophage inflammatory proteins 1alpha and 1beta as a native secreted heterodimer. J Biol. Chem. 276(15), 12404-12409 (2001).

14. Dudek AZ, Nesmelova I, Mayo K, Verfaillie CM, Pitchford S, Slungaard A. Platelet factor 4 promotes adhesion of hematopoietic progenitor cells and binds IL-8: novel mechanisms for modulation of hematopoiesis. Blood. 101(12), 4687-4694 (2003).

15. Carlson J, Baxter SA, Dreau D, Nesmelova IV. The heterodimerization of platelet-derived chemokines. Biochim. Biophys. Acta. 1834(1), 158-168 (2013).

16. Paoletti S, Petkovic V, Sebastiani S, Danelon MG, Uguccioni M, Gerber BO. A rich chemokine environment strongly enhances leukocyte migration and activities. Blood. 105(9), 3405-3412 (2005).

17. Hoogewerf AJ, Kuschert GS, Proudfoot AE, Borlat F, Clark-Lewis I, Power CA, Wells TN. Glycosaminoglycans mediate cell surface oligomerization of chemokines. Biochemistry. 36(44), 13570-13578 (1997).

18. Lau EK, Paavola CD, Johnson Z, Gaudry JP, Geretti E, Borlat F, Kungl AJ, Proudfoot AE, Handel TM. Identification of the glycosaminoglycan binding site of the CC chemokine, MCP-1: implications for structure and function in vivo. J Biol. Chem. 279(21), 22294-22305 (2004).

19. Proudfoot AE, Handel TM, Johnson Z, Lau EK, LiWang P, Clark-Lewis I, Borlat F, Wells TN, Kosco-Vilbois MH. Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. Proc. Natl. Acad. Sci. U. S. A. 100(4), 1885-1890 (2003).

20. Kuschert GS, Hoogewerf AJ, Proudfoot AE, Chung CW, Cooke RM, Hubbard RE, Wells TN, Sanderson PN. Identification of a glycosaminoglycan binding surface on human interleukin-8. Biochemistry. 37(32), 11193-11201 (1998).

21. Poluri KM, Joseph PR, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. J. Biol. Chem. 288(35), 25143-25153 (2013).

22. Rajasekaran D, Keeler C, Syed MA, Jones MC, Harrison JK, Wu D, Bhandari V, Hodsdon ME, Lolis EJ. A model of GAG/MIP-2/CXCR2 interfaces and its functional effects. Biochemistry. 51(28), 5642-5654 (2012).

23. Rajarathnam K, Kay CM, Dewald B, Wolf M, Baggiolini M, Clark-Lewis I, Sykes BD. Neutrophil-activating peptide-2 and melanoma growth-stimulatory activity are functional as monomers for neutrophil activation. J Biol. Chem. 272(3), 1725-1729 (1997).

24. Shao W, Jerva LF, West J, Lolis E, Schweitzer BI. Solution structure of murine macrophage inflammatory protein-2. Biochemistry. 37(23), 8303-8313 (1998).

25. Payne JW. Polymerization of proteins with glutaraldehyde. Soluble molecular-weight markers. Biochem. J. 135(4), 867-873 (1973).

26. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods. 9(7), 671-675 (2012).

27. Vehlow C, Stehr H, Winkelmann M, Duarte JM, Petzold L, Dinse J, Lappe M. CMView: interactive contact map visualization and analysis. Bioinformatics. 27(11), 1573-1574 (2011).

28. Rochus Keller. The Computer Aided Resonance Assignment Tutorial.ISBN- 3-85600-112-3 (1966).

29. Smith DF, Galkina E, Ley K, Huo Y. GRO family chemokines are specialized for monocyte arrest from flow. Am. J Physiol Heart Circ. Physiol. 289(5), H1976-H1984 (2005).

30. Luttichau HR. The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist. J Biol. Chem. 285(12), 9137-9146 (2010).

31. Haskill S, Peace A, Morris J, Sporn SA, Anisowicz A, Lee SW, Smith T, Martin G, Ralph P, Sager R. Identification of three related human GRO genes encoding cytokine functions. Proc. Natl. Acad. Sci. U. S. A. 87(19), 7732-7736 (1990).

32. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, Rousseau S, Eidelman DH, Hamid Q. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. J Immunol. 191(5), 2731-2741 (2013).

Chapter 4: Biophysical Characterization of CXCL3 and its Comparison with CXCL2

Abstract

CXCL3 is the third duplicated member of GRO chemokines. Although structural details are available for CXCL1 and CXCL2, no such information regarding CXCL3 is available till date. In the present chapter, CXCL3 was cloned, purified, and characterized using NMR and other biophysical techniques. Biophysical studies revealed that overall structure, oligomerization, thermal stability, and heparin binding characteristics of CXCL3 are similar to CXCL2. Tertiary structural characteristics obtained using surface electrostatic potentials and ANS fluorescence revealed that the closest paralogs CXCL2 and CXCL3 have distinct charge distribution on the quaternary surface, and a specific hydrophobic pocket in CXCL3. Further, residue level structural stabilities and conformational dynamics established the lower stability and enhanced dynamic nature of CXCL3 in contrast to CXCL2. All these studies delineate the differential structure-stability-function relationships of GRO proteins.

4.1 Introduction

CXCL3 (GRO γ), is a member of GRO family of chemokines [1]. As discussed in **Chapter 2**, GRO family is a resultant of gene duplication event [2-6]. Hence the GRO members are closely related to each other, and are also involved in numerous common functions [7]. However, multitude of biological studies have also reported distinctive functions performed by the GRO genes that are regulated in tissue and signal specific manner [8-10]. Comparative CXCR2 chemotactic activity studies of GRO genes evidenced highest efficacy for CXCL1 and intermediate efficacies for CXCL2 and CXCL3 [9]. Differential oligomerization features for CXCL1 and CXCL2 discerned in **Chapter 3**, also reflects towards differential regulatory patterns followed by GRO chemokines. It has also been shown that CXCL3 is involved in diverse variety of specific functions, and follows distinctive molecular mechanism to accomplish its functions as

compared to other GRO chemokines [11,12]. Altogether, these studies suggest that the newly formed GRO genes have acquired novel functions and specialized mechanisms to meet the system requirements in addition to the performance of their basal GRO chemokine functions.

Albeit structural details, oligomerization patterns, receptor/GAG binding interactions are elucidated for both CXCL1 and CXCL2 chemokines in **Chapter 3** and elsewhere [13-16]. In this chapter, residue level structural, dynamic, and stability studies of CXCL3 were done by using a recombinant CXCL3 that was cloned and purified in a pET-32 bacterial expression vector. All the structural and stability parameters were compared with its closest paralog CXCL2.

4.2 Materials and methods

4.2.1 Cloning of CXCL3

Full length CXCL3 gene (NCBI Ref seq: NM_203320.2) was purchased from Sino-biologicals, China in pMD18-T cloning vector (Cat. No. MG50258-M). The gene encoding the CXCL3 protein was amplified using the appropriate primers:

CXCL3_FP: 5'GCAAGGTACCGAAAACCTGTATTTTCAGGGGGCTGTTGTGGCCAGTGAGCT-3'
CXCL3_RP: 5'- GCAACCATGGTCAGCTGGACTTGCCGCT- 3'

The amplified gene was sub-cloned into a pET32 bacterial expression vector between the Kpn1 and Nco1 restriction sites. The gene insertion was confirmed from the gene sequencing results. The cloning of CXCL3 was carried out using the method described earlier in **Section 3.2.1 (Chapter 3)**.

4.2.2 Expression, purification, and in-vitro folding of CXCL3

CXCL3 was expressed using the same protocol as described in **Section 3.2.2 (Chapter 3).** CXCL3 expression was checked at two different conditions after induction in which the cells were grown at (a) 37 ºC for 4 hrs and (b) 20 ºC for 18 hrs. Expression of CXCL3 in both supernatant and pellet was analyzed on 12 % SDSPAGE. Large scale expression of CXCL3 protein was carried out as described in **Section 3.2.2 (Chapter 3)**.

Protein (natively folded CXCL3/CXCL3_NF) from the soluble fraction/supernatant was purified using the protocol as described in **Section 3.2.2 (Chapter 3)**. Protein present in the

insoluble fraction (inclusion bodies) was solubilized/denatured at room temperature in 8 M urea lysis buffer containing 20 mM Tris (pH 8) and 500 mM NaCl, using the cell homogenizer at room temperature. Lysate was then centrifuged at 14000 rpm for 20 min, at 20 ºC. Supernatant containing the denatured protein was separated from the pellet containing the cell debris. Denatured protein was refolded by dialyzing it twice (8 hrs for each cycle of dialysis) in the buffer containing 20 mM Tris (pH 8) and 500 mM NaCl. The refolded protein (CXCL3_RF) was obtained by centrifugation. The supernatant fraction containing the CXCL3 fusion protein was purified using the same protocol as described in **Section 3.2.2 (Chapter 3)**.

4.2.3 Size exclusion chromatography

Natively folded (NF), and refolded (RF) CXCL3 proteins were subjected to SEC using the as described in **Section 3.2.3 (Chapter 3)**.

4.2.4 Heparin binding assay

Heparin binding assay was performed using 1 ml HiTrap Heparin high performance column (GE) using GE-AKTA prime. The column was pre-equilibrated with 20 mM Tris (pH 7), 50 mM NaCl. Proteins CXCL1, CXCL2, and CXCL3, at fixed concentration of 0.1 mg/ml were injected in the column, and were eluted using a linear gradient of 0 M to 2 M NaCl at a flow rate of 1 ml/min by monitoring the absorbance at 215 nm.

4.2.5 Circular dichroism (CD) spectroscopy

Far UV-CD experiments were carried out on Jasco J-1500 CD spectrophotometer using 1 mm path length quartz cuvette in the range of 190-250 nm. CXCL2 and CXCL3 samples concentrated to 50 µM in 20 mM Tris (pH 7), 50 mM NaCl buffer were used for CD experiments. Thermal denaturation of CXCL2 and CXCL3 was monitored in the temperature ranges 20 ºC to 90 ºC by recording the spectra at regular intervals of 10 ºC with the gradual increase in the temperature (1 ºC/min) and an incubation time of 5 min at each resting temperature. To analyze the reversibility of structural transitions, proteins were cooled back to 20 ºC from 90 ºC with gradual decrease in the temperature (1 ºC/min) at regular intervals and the spectra were recorded again at 20 ºC.

4.2.6 Fluorescence spectroscopy

All the fluorescence experiments were carried out on Fluorolog-3 spectrofluorimeter (Horiba Jobin Yvon Spex®) containing xenon lamp source at 25 ºC using 4 mm path length quartz cuvette with 2.5 nm slit widths. To monitor the 8-Anilinonaphthalene-1-sulfonic acid (ANS) binding, CXCL2 and CXCL3 samples were concentrated to 50 µM in 20 mM Tris (pH 7), 50 mM NaCl buffer. Samples were excited at 380 nm and the emission spectra were scanned in the range of 400-600 nm. A fixed concentration of 1:5 (protein: ANS) was used to record the CXCL2 and CXCL3 spectra.

Job plot

Job plot analysis (continuous variation analysis) was carried out by recording the ANS fluorescence at a fixed total protein+ANS concentration of 100 µM, while varying their relative mole fractions. Fluorescence maxima (F_{max}) was obtained from each spectrum and difference in F_{max} of bound ANS and F_{max} of free ANS was plotted against the mole fraction of ANS. The mole fraction of ANS at point of inflection was used to calculate the ANS protein binding stoichiometry (n) according to the simple equation:

$$n = 1- \chi_{ANS} / \chi_{ANS} \qquad \textbf{(4.1)}$$

Where, χ_{ANS} is the mole fraction of ANS at point of inflection.

ANS fluorescence titration

ANS fluorescence titration was carried out using fixed ANS concentration and varying the concentration of CXCL3 in increasing order. 100 µM of ANS was titrated with 100 µM CXCL3, starting with the initial concentration of 1 µM of protein. Fluorescence emission at each concentration of protein was recorded by scanning in the range of 400-600 nm. Fluorescence emission maxima (F_{max} (bound + unbound)) obtained at different concentrations of protein (1, 2, 4, 6, 8, 10, 15, 20, 40, 60, 80, 100 (in µM)) was subtracted from the fluorescence maxima of free ANS (F_{max} free), thus yielding F_{max} for the bound fraction (F_b), which corresponds to fluorescence purely from protein:ANS complex.

$$F_b = F_{max}\,(\text{bound} + \text{unbound}) - F_{max}\,(\text{free}) \qquad \textbf{(4.2)}$$

Fluorescence intensities (F_b) were plotted against their corresponding protein concentrations in order to determine the binding constant; the data was fitted to the following equation:

$$F= F_b [L]/ K_d + [L] \quad \textbf{(4.3)}$$

Where, F is the fluorescence intensity at concentration L of the protein, F_b is the maximum fluorescence intensity that can be obtained, and K_d is the binding constant, which is equal to the concentration at which 50% of the protein is bound to the ANS.

Fluorescence life time measurement

Fluorescence life time measurements were carried out using the time correlated single photon counting spectrometer provided by Horiba Jobin Yvon (Fluorocube fluoresence life time system with Nano LED). Samples (ANS alone, ANS:CXCL2 and ANS:CXCL3) were excited at 380 nm using Nano LED in a 1cm quartz cell at 25 ºC. Fluorescence emission was monitored at wavelength 495 nm for ANS alone, ANS: CXCL2, and ANS: CXCL3 using TBX-04-D detector. Decay traces were measured using 2048 channel analyzer. Decay curves were analyzed using DAS 6.1 (Data Station) provided by IBH and was fitted by multi-exponential iterative reconvolution technique to obtain the best chi square value ($\chi 2$) with minimum standard deviations.

4.2.7 NMR spectroscopy

NMR samples

For NMR studies, ^{15}N and $^{13}C/^{15}N$ labeled CXCL3 proteins were exchanged with 50 mM sodium phosphate buffer, pH 6.0, with 0.01 % sodium azide. Final volume of all the NMR samples was 550 µl with 90 % H_2O and 10 % D_2O. 1 mM samples of ^{15}N- CXCL3_NF and CXCL3_RF were used for ^{1}H-^{15}N HSQC experiment. 1mM $^{13}C/^{15}N$ CXCL3 sample was used for all 3D NMR experiments. Relaxation and temperature dependence experiments were carried out using 750 µM of ^{15}N-CXCL2 and ^{15}N-CXCL3. For hydrogen exchange, DOSY, and ANS titration experiments, ^{15}N-CXCL2 and ^{15}N-CXCL3 were concentrated to 200 µM, but for hydrogen exchange, samples were lyophilized and re-dissolved in 100 % D_2O.

NMR data acquisition and processing

All the NMR experiments were carried out using a triple channel Bruker 500 MHz/800 MHz spectrometer equipped with a TXI cryoprobe at 25 ºC. ^{1}H and ^{15}N carrier frequencies were set to 4.7 ppm and 120.5 ppm respectively in all the experiments.

All ^{1}H-^{15}N HSQC spectra were recorded with 32 scans and 256 increments in the indirect dimension with spectral width 16 ppm and 28 ppm for ^{1}H and ^{15}N dimensions respectively. For the backbone resonance assignments, standard three dimensional NMR experiments including HNCA, HNCACB, HNCO, CBCACONH, TOCSY, and NOESY were carried out on Bruker 800 MHz using $^{13}C/^{15}N$-CXCL3 (1 mM). For 3D-NMR experiments, ^{13}C carrier frequency was set to 53 ppm for HNCA; 42 ppm for HNCACB, CBCA(CO)NH; at 173 ppm for HNCO.

Translational diffusion NMR

2D-DOSY experiments were carried out on Bruker 800 MHz spectrometer with an actively shielded pulse field gradient unit with a power of producing maximum gradient with the strength of 53 G/cm. DOSY spectra were acquired by employing the stimulated echo bipolar gradient pulse program "ledbpgppr2s" (encompassing longitudinal eddy-current delay block) by using diffusion delay (d20) of 300 ms and pulse field gradients (P30) whose amplitudes varying from 5-95 % in 16 equally spaced steps, with duration of 3 ms and recovery delay (D16) of 100 µs for each step. Experiment was recorded with 64 scans with recycle delay (D1) of 2 s. Spectra were subjected to the usual Fourier transformation and baseline correction, followed by DOSY processing using eddosy" and "dosy2d" analysis modules of the Bruker Topspin software. All the DOSY measurements were carried out in 100 % D_2O solvent. The proteins CXCL2, hen egg lysozyme (HEL) and chicken SH3 domain were used as reference molecular weight markers to assess the molecular weight of CXCL3.

ANS titration experiment

^{15}N-CXCL2 and ^{15}N-CXCL3 were titrated with ANS at increasing ANS: Protein molar ratios including 0.1, 0.2, 0.3, 0.5, 0.75, 1, 2, and 4. ^{1}H-^{15}N HSQC spectra were recorded for each

molar ratio sample at 500 MH_Z. residue wise chemical shift perturbations (CSP) at molar ratio of 2 (protein:ANS= 1:2) were calculated for CXCL3 using the following equation:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2} \quad \textbf{(4.4)}$$

Where, $\Delta\delta_H$ and $\Delta\delta_N$ are the change in chemical shift of proton and nitrogen respectively.

Relaxation NMR

The complete set of relaxation experiments including T_1, T_2, and ^1H-^{15}N NOE were recorded for ^{15}N-CXCL2 and ^{15}N-CXCL3 on the Bruker 800 MHz spectrometer. The relaxation measurements were carried out using the pulse sequences described by Peng and Wagner, and Farrow et al [17,18]. Both R_1, R_2 relaxations were collected using a delay of 2 s. ^{15}N longitudinal relaxation (T_1/R_1) was carried out using following relaxation delays: 10, 50, 100, 200, 300, 400, 500, and 600 ms. For ^{15}N transverse relaxation rates (T_2/R_2), the following relaxation delays were employed: 15.36, 30.72, 46.08, 61.68, 77.04, 92.16, 107.52 ms. R_1 and R_2 values were calculated for each cross peak by fitting the measured peak intensities to single exponential decay equation:

$$I(t) = I_0 e^{-R_{1,2} t} \quad (4.5)$$

Where, I and I_0 are the intensities at time t and 0; $R_{1,2}$ t is the relaxation rate R_1 or R_2; t is the time. ^1H-^{15}N steady state NOE experiments with and without proton saturation were carried out using proton saturation time of 2.5 sec and a relaxation delay of 2.5 s. Steady state ^1H-^{15}N NOEs were calculated as the ratio of peak intensities with and without proton saturation. The errors in the NOEs were obtained using the root mean square value of the background noise as described by Farrow et al [18].

Hydrogen exchange

Theory

Amide protons in a fully structured protein can be present either in the structural part of the protein (not accessible to the outside environment) or in the solvent accessible regions. This indicates the presence of hydrogen's in a protected state (non- exchangeable state) or in a transiently open state (exchangeable state) [19].

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} NH_{open} \xrightarrow{k_{rc}} Exchanged \quad \textbf{(4.6)}$$

where, k_{op} is opening rate, k_{cl} is the closing rate, and k_{rc} is the intrinsic exchange rate of amide protons. The k_{obs} can be calculated from the above equation (4.6) under steady state conditions, which is given by eq. (4.7):

$$k_{obs} = \frac{k_{op}\, k_{rc}}{k_{op} + k_{cl} + k_{rc}} \quad \textbf{(4.7)}$$

Equation 4.7 can be rewritten as equation 4.8, as under the native conditions, $k_{op} << k_{cl}$,

$$k_{obs} = \frac{k_{op}\, k_{rc}}{k_{op} + k_{cl}} \quad \textbf{(4.8)}$$

Under bimolecular exchange conditions, such as under low pH and temperature conditions, values of $k_{cl} >> k_{op}$, equation 4.8, can be simplified and written as

$$k_{obs} = k_{rc}\left(\frac{k_{op}}{k_{cl}}\right) \quad \textbf{(4.9)}$$

Therefore, the free energy of stabilization can be calculated as

$$\Delta G_{HX} = -RT\, ln\left(\frac{k_{op}}{k_{cl}}\right) \quad \textbf{(4.10)}$$

$$\Delta G_{HX} = -RT\, ln\, K_{op} \quad \textbf{(4.11)}$$

Where, R is the gas constant and T is the absolute temperature, and K_{op} is equilibrium constant of unfolding.

HX data acquisition

HX experiments were recorded for the both the CXCL2 and CXCL3 on Bruker 800 MHz equipped with a TXI cryo-probe. The dead time of 6 minutes was given, which is an incubation time given to the protein after adding D_2O and before the start of first spectra. Samples were loaded on pre-shimmed and pre-tuned NMR spectrometer. Each HSQC experiment was recorded for the period of 20 minutes (16 scans and 128 complex increments in the indirect ^{15}N dimension). The experiments were recorded for 40 hours to monitor the decay of NH intensities.

HX analysis

Data was processed in Bruker Topspin 3.5, and the data was analyzed in sigma plot. This decay profile of each residue was fitted to single exponential curve to yield k_{obs}, which was used to calculate Kop (= k_{obs}/k_{rc}), where krc is the theoretically calculated values for the residues in the specific tripeptide by Bai et al [20]. Free energy of HX (ΔG_{HX}) was calculated according to equation **4.11** by considering the EX2 mechanism at pH 6.0.

Temperature dependence

^{1}H-^{15}N HSQC spectra at different temperatures including 285 K, 290 K, 295 K, 300 K, 310 K, 315 K, 320 K, were recorded on Bruker 800 MHz for CXCL2 and CXCL3. Amide proton chemical shifts as a function of temperature in the range 285-320 K were obtained and their temperature coefficients were calculated by fitting the proton chemical shifts to a linear equation. To analyze the curved temperature dependence of each residue, the residual values were calculated by deriving the deviations of experimental chemical shift at each temperature [21].

4.2.8 In-silico structural analysis.

Structure for CXCL2 (PDB ID: 1MI2, 3N52) was available in RCSB protein data bank. Structural models for CXCL1 and CXCL3 were generated as described in **Section 2.2.9 (Chapter 2)**. Structures of CXCL1 and CXCL2 were aligned with CXCL3 in PyMol using 'align' macro and the RMSD values were calculated. [22]. Comparative contact maps of CXCL3 with CXCL1 and CXCL2 were generated using the method described in **Section 3.2.5 (Chapter 3)**.

4.2.9 Molecular docking

Molecular docking approach was used to investigate the binding site of ANS to CXCL3. ANS was docked on the modeled structure of CXCL3 using default parameter set of Autodock Vina [23]. Hydrogen atoms were added and Kollman charges (3.0) were assigned to all the atoms of the protein. The grid was defined on the basis of NMR chemical shift constraints obtained from chemical shift perturbation experiments. The grid was centered at (25.542, 3.624, 33.756) of dimension 30*30*30 with 0.314 Å spacing.

4.3 Results

4.3.1 Cloning and expression of CXCL3

The amplified CXCL3 gene from pMD18-T cloning vector has been successfully ligated using T4 ligase between the Kpn1 and Nco1 restriction sites of pET32 expression vector. The vector contains thioredoxin (Trx) protein as a fusion moiety to enhance the solubilization of protein of interest, and a hexa histidine affinity tag that aids in facile purification of the fusion protein. A TEV protease cleavage sequence (ENLYFQG) was inserted between the fusion protein and the gene of interest in order to facilitate the removal of tags. The amplified product of CXCL3 gene has been shown in the **Fig. 4.1 A**. The successful cloning of the CXCL3 gene into pET32 has been confirmed by DNA sequencing result **(Fig. 4.1 B)**.

In order to check the expression of the CXCL3 gene in pET32 vector, *E. coli* (BL21) cells transformed with recombinant CXCL3 plasmid and induced with IPTG, and the post induced cultures were grown at two different temperatures (37 ºC and 20 ºC). Although at both the temperatures, CXCL3 protein has been expressed, it showed a better expression profile at 20 ºC compared to 37 ºC **(Fig. 4.1 C)**. Soluble and insoluble fractions obtained from the cell lysates from both the cultures (grown at 37 ºC and 20 ºC) were analyzed on 12% SDS-PAGE gel by observing the over expressed band of Trx-CXCL3 (~25 kD) in SDS-PAGE **(Fig. 4.1 C)**. PAGE analysis evidenced that the cells grown at 37 ºC expressed CXCL3 exclusively in inclusion bodies (insoluble form) as misfolded protein. Whereas, the cells grown at 20 ºC cells, expressed CXCL3 partly soluble form (supernatant) in cytoplasm (40 %) and rest as misfolded protein in the insoluble fraction in inclusion bodies (60 %) **(Fig. 4.1 C)**.

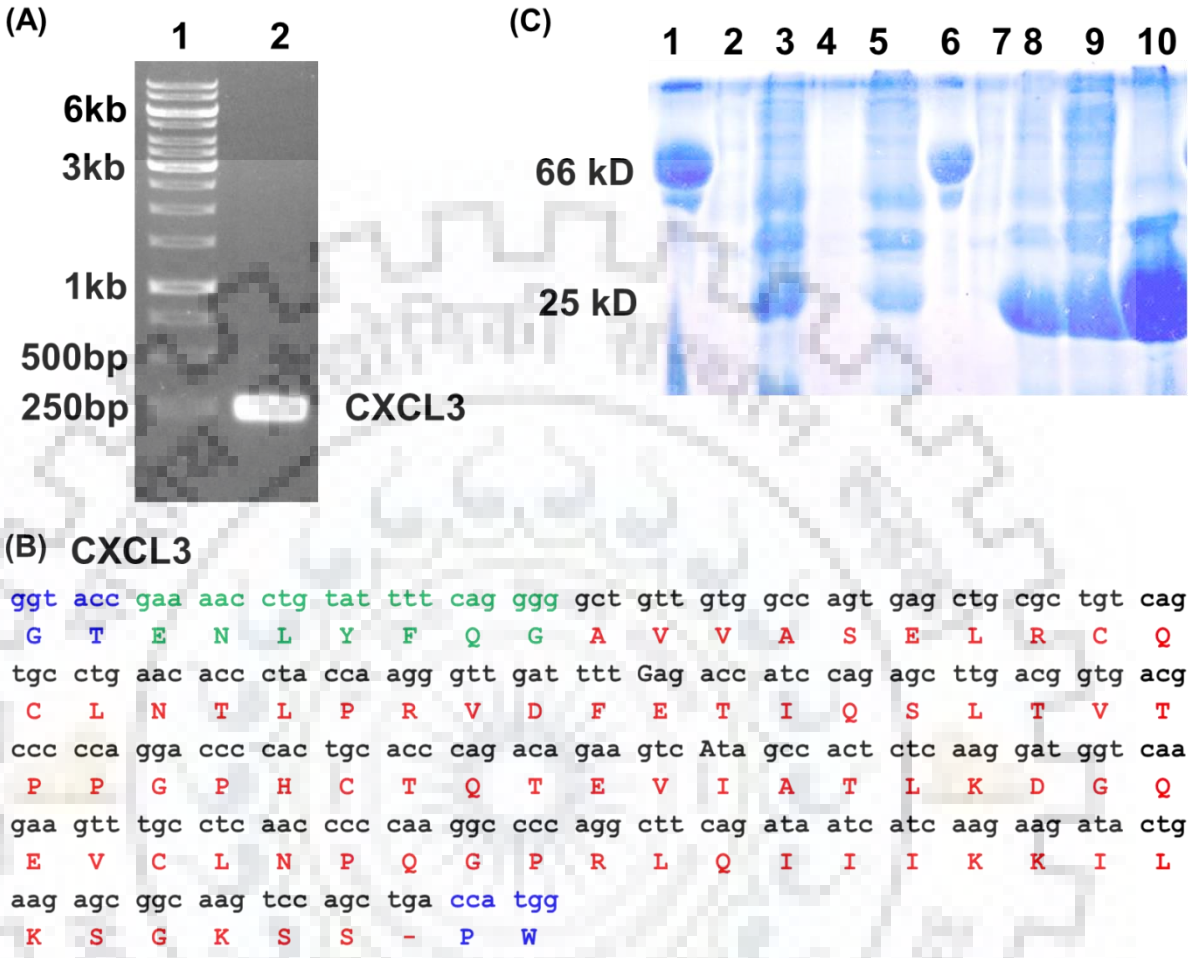


Figure 4.1: (A) *1.5 % agarose gel showing the 1 kB DNA ladder in lane 1, and an amplified CXCL3 gene (255 base pairs) in lane 2.* **(B)** *Sequencing result (255 base pairs) confirming the presence of CXCL3 nucleotide sequence (black) and protein sequence (red) along with TEV cleavage site (green) and restriction sites Kpn1 at N-terminus and Nco1 at C-terminus (blue). The amino acid sequence for the encoded gene is presented with single letter code.* **(C)** *12 % SDS-PAGE analysis of expression of CXCL3 at 20 ºC, Lanes 1, 6, represent marker, Lanes 2, 3, 4, 5 represent un-induced, induced CXCL3, soluble fraction, and insoluble fractions obtained after cell lysis from the cells grown at 37 ºC, 4 hours, Lanes 7, 8, 9, 10, represent un-induced, induced CXCL3, soluble fraction, and insoluble fractions obtained after cell lysis from the cells grown at 20 ºC, 18 hours.*

4.3.2 Purification of CXCL3 chemokine

To maximize the protein yield, CXCL3 protein was purified from both the soluble and insoluble fractions. The overall purification protocol of CXCL3 is outlined in **Fig. 4.2**. Prior to the

purification of CXCL3 contained in the insoluble fraction, it was subjected to solubilization steps that include; (1) denaturation using urea and (2) refolding through dialyzing the urea **(Fig. 4.2)**. After refolding, CXCL3_RF (Refolded CXCL3) is subjected to the same purification protocol as employed for the CXCL3_NF (Natively folded CXCL3). A combination of chromatography techniques (affinity, ion exchange and SEC) was applied to purify the CXCL3 protein. The purity at every step of the purification process was assessed by SDS-PAGE, and the yields were determined from BCA assay **(Fig. 4.3,** and **Table 4.1)**.

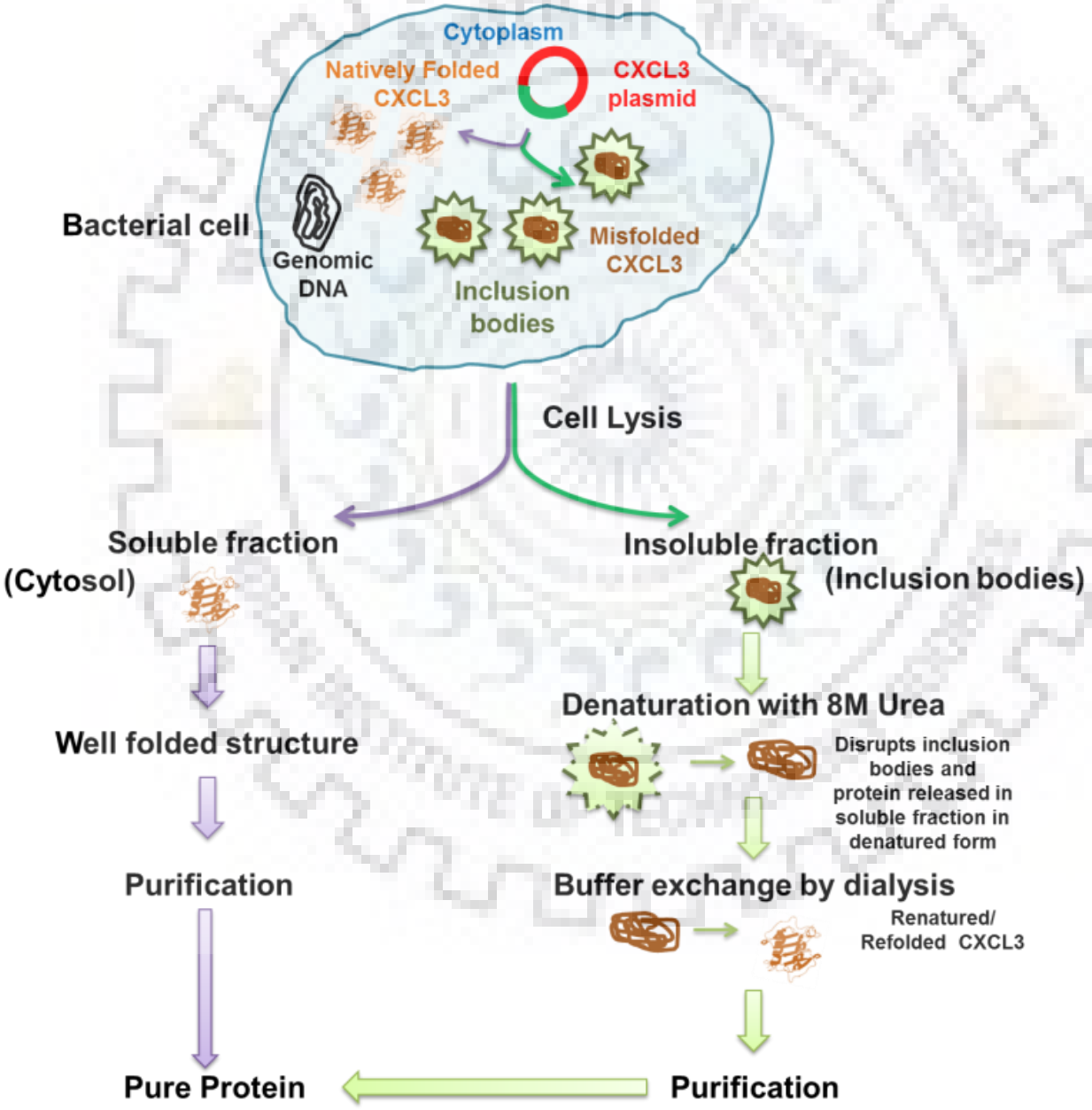


Figure 4.2: *Schematic showing the purification protocol of CXCL3.*

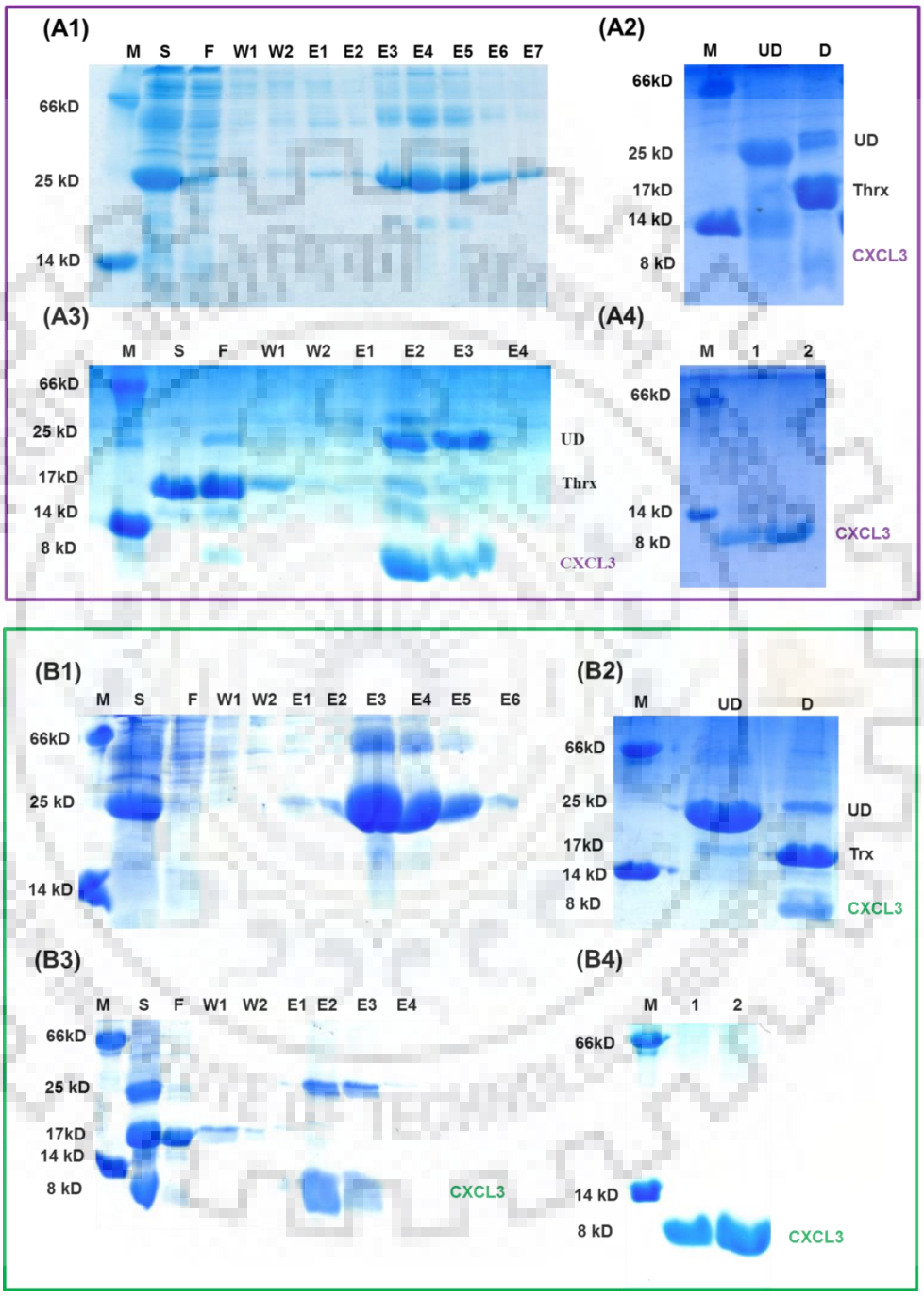


Figure 4.3: *15 % SDS-PAGE gel analysis of protein migration at different purification steps for CXCL3_NF (upper purple panel) and CXCL3_RF (lower green panel).* **(A1/B1)** *1st Ni NTA purification step: Lane S- bacterial cell lysate, Lane F-flow through fraction, Lanes W1 and W2 - washes with 10 mM and 30 mM imidazole, Lanes E1, E2, E3, E4, E5, E6 - eluted fractions using*

400 mM imidazole. **(A2/B2)** *TEV Digestion step: Lane UD– undigested sample of 25 kD, Lane D is the digested sample with three bands - the left over undigested sample (25 kD), thioredoxin tag (17 kD), and CXCL3 protein (8 kD);* **(A3/B3)** *Trx fusion tag separation step using cation exchange chromatography: Lane S; digested sample as supernatant, Lane F- flow through; Lanes W1 and W2- washes with 50 mM NaCl and 100 mM NaCl, Lanes E1, E2, E3, E4 are eluted fractions using 500 mM NaCl.* **(A4/B4)** *2nd Ni-NTA chromatography purification step: Lanes 1, 2, and 3- flow through fractions containing 8 kD CXCL3 protein. In all the gels, Lane M- marker containing two standard proteins lysozyme (14 kD) and BSA (66 kD).*

The purification protocol involved four steps: (a) In the first step, affinity chromatography was performed using Ni-NTA to bind the fusion protein using its $(His)_6$ Tag. Most of the non-specific impurities were washed with a gradient of imidazole and the fusion protein (~ 25 kD) was eluted in various fractions **(Fig. 4.3 A1/B1)**. (b) In the second step, the fractions containing the fusion protein were pooled, twice dialyzed to remove the imidazole, and the resultant protein was subjected to TEV enzyme cleavage. TEV digestion resulted in three fractions; a small amount of uncut fused protein (~ 25 kD), Trx-His tag (~ 17 kD) and the CXCL3 protein (~ 8 kD) **(Fig. 4.3 A2/B2)**. (c) The CXCL3 protein (pI – 8.5) was then separated from fusion tag (pI – 5.1) using cation exchanger (SP-sepharose). The elution profiles of CXCL3 using 500 mM NaCl contain mainly CXCL3 protein and some amount of the fused protein **(Fig. 4.3 A3/B3)**. (d) CXCL3 was then separated from the undigested fusion protein using a 2nd Ni-NTA step. As CXCL3 does not have a tag on its own, it will not bind to Ni-NTA and has been collected in flow through **(Fig. 4.3 A4/B4)**.

Table 4.1*: Represents the amount of natively folded and refolded CXCL3 protein obtained from 1 L of LB culture after each step of purification.*

Purification step	NF_CXCL3 (mg)	RF_CXCL3 (mg)
1st Ni-NTA purification	80 ± 5	120 ± 10
Dialysis	63 ± 5	83 ± 5
Cation exchange chromatography	22 ± 2	28 ±2
Reverse Ni NTA purification	8 ± 2	10 ± 2
Gel filtration chromatography	6.5 ±1	8.5 ± 1
Total recombinant CXCL3 obtained	**15 ± 1 mg/L**	

CXCL3 obtained from the soluble (CXCL3_NF) and insoluble fractions (CXCL3_RF) were subjected to size exclusion chromatography and elution profiles were compared **(Fig. 4.4 inset)**. The protein obtained from both the fractions was eluted at the same volume (~ 72 ml), indicating the presence of same conformation/oligomeric state in both refolded and natively folded CXCL3. An overall yield of ~ 15 mg of pure recombinant protein has been obtained by exploiting the Trx fusion protein expression and purification protocol in a bacterial expression system **(Table 4.1)**.

Further, ^{1}H-^{15}N HSQC spectra for both CXCL3_NF and CXCL3_RF was recorded using ^{15}N labeled CXCL3 samples to confirm the native state conformation of the recombinantly produced proteins **(Fig. 4.4)**. Overlaying the spectrum for both natively soluble and refolded proteins indicated that both the spectra are completely identical. Moreover, the ^{1}H-^{15}N HSQC spectra showed an excellent peak dispersion and contained ~ 70 (total amino acids in CXCL3 - 73) well resolved peaks indicating that under the chosen concentration, the protein exist in a single natively folded conformation. Comparison of the CXCL3 ^{1}H-^{15}N HSQC spectrum with the reported CXCL1/CXCL2 NMR spectra from murine **(Fig. 3.6, Chapter 3)**, and human established that recombinant CXCL3 protein adapted a native CXC chemokine fold [13,24-26].

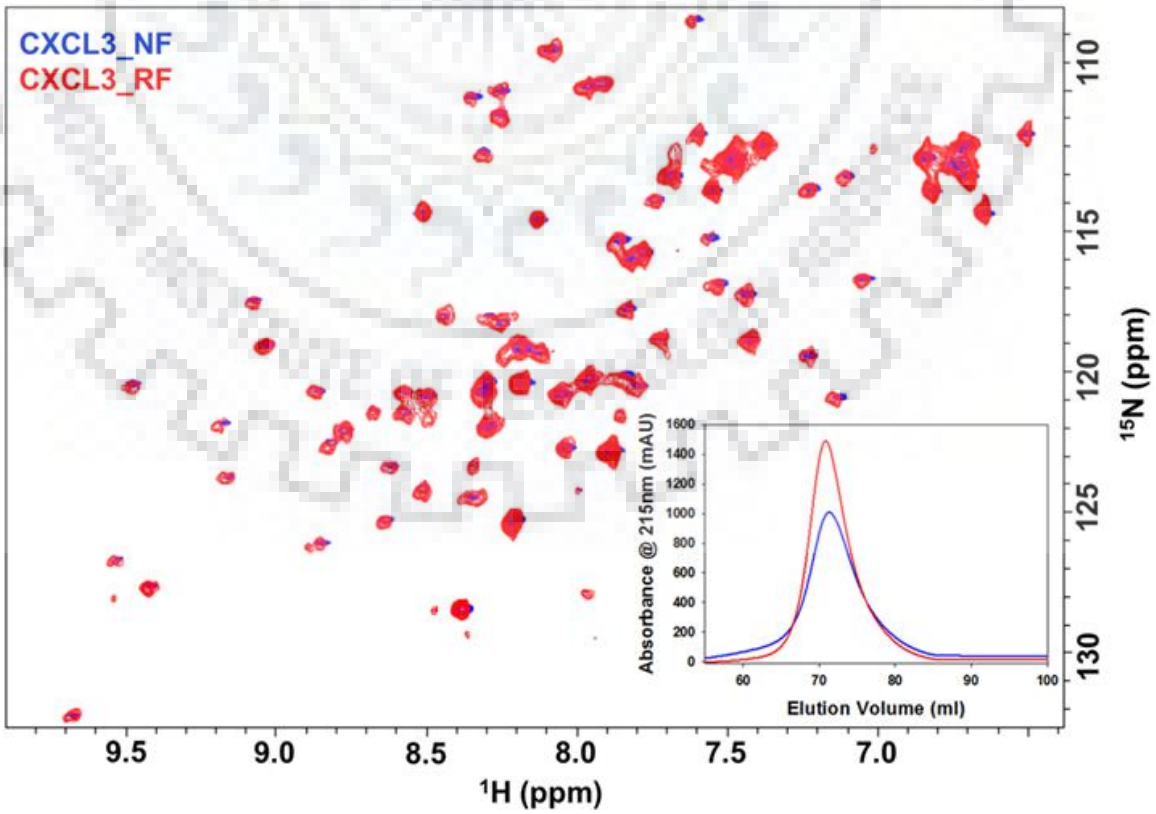


Figure 4.4: *Overlay of ^{1}H -^{15}N HSQC spectra of CXCL3_NF (blue) and CXCL3_RF (red) at 25 ºC. The size exclusion chromatography elution profiles of CXCL3_NF and CXCL3_RF are presented in the inset.*

4.3.3 Resonance assignment of CXCL3

In order to characterize the structural features of CXCL3, NH cross peaks in its HSQC spectrum were assigned using conventional triple resonance experiments as described in **Section 4.2.7**. Out of 73 residues, 6 were proline. All the NH resonances corresponding to the remaining 67 residues were assigned. The HNCACB strips for the stretch V40-L44 are shown in **Fig. 4.5**. Backbone NH assignments for the assigned residues are marked in **Fig. 4.6**.

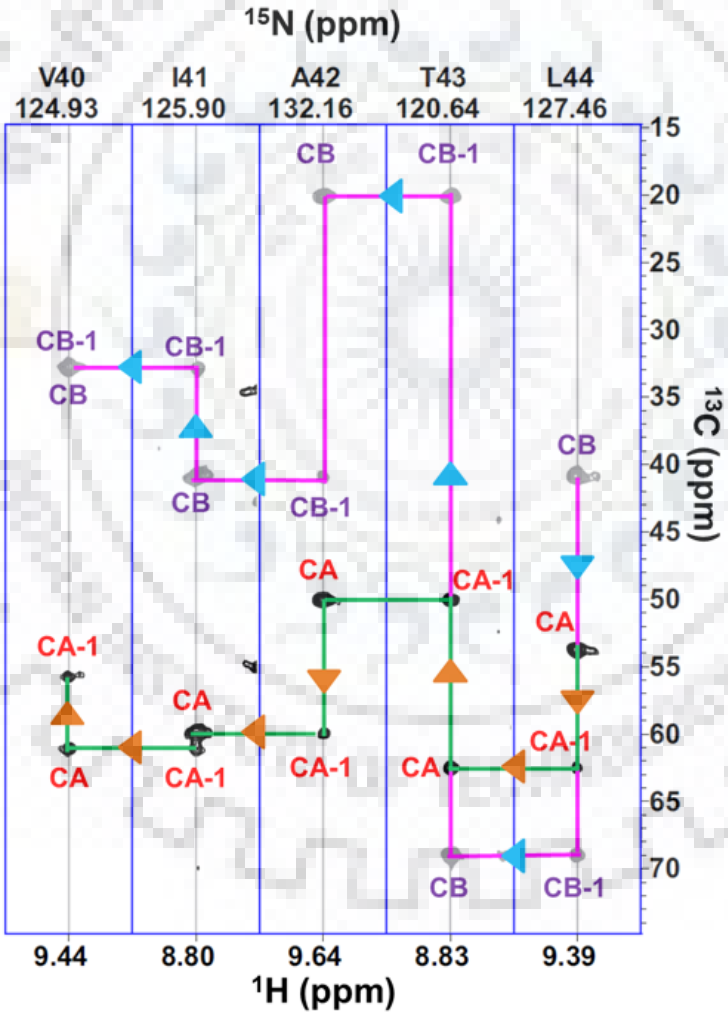


Figure 4.5: *HNCACB strip plot for a stretch of five residues (V40-I41-A42-T43-L44) showing the connectivities of C_α resonances with pink lines and blue arrow heads and of C_β resonances with green lines and yellow arrow heads for CXCL3.*

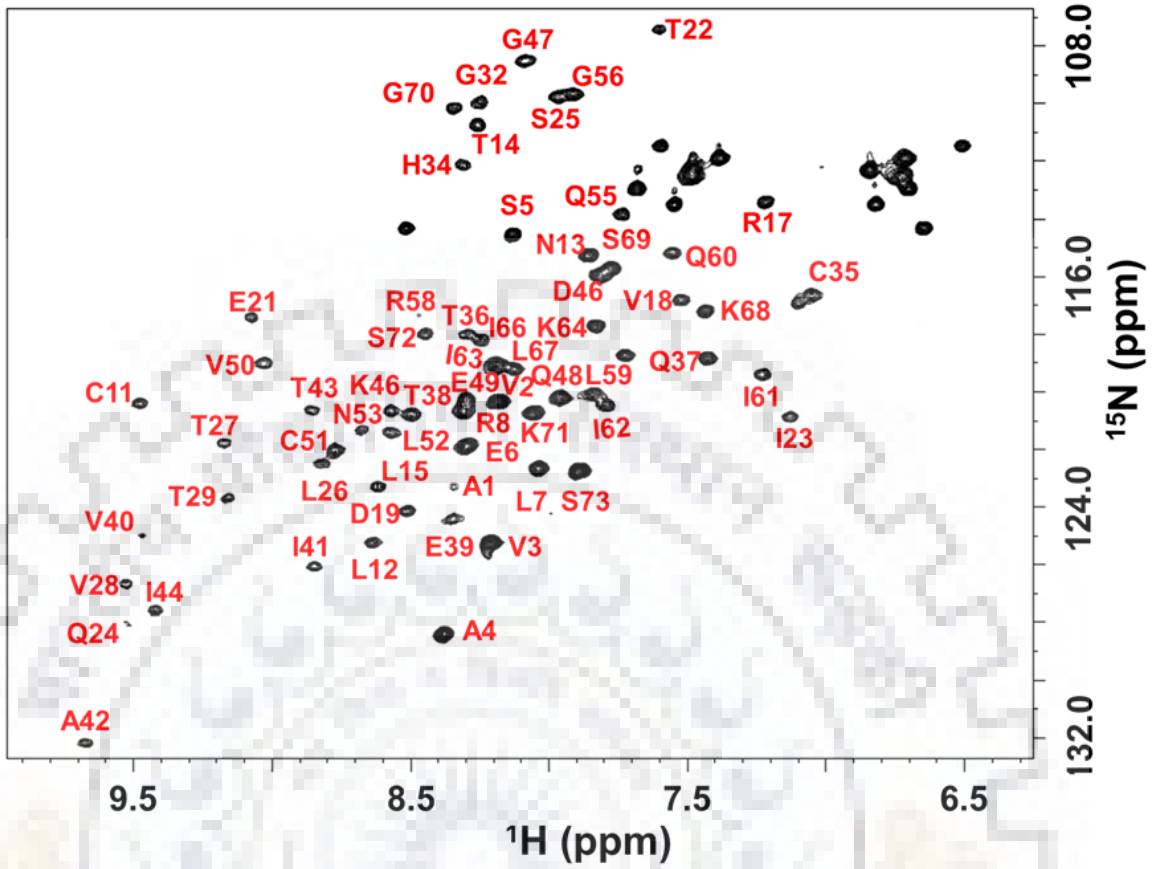


Figure 4.6: *^{1}H-^{15}N HSQC spectrum showing the residue specific resonance assignments for CXCL3. The NH cross peaks are labeled with residue symbol and number.*

4.3.4 Characterizing the oligomeric state of CXCL3

Assessing the oligomerization state of the CXCL3 in the observed conformation is essential as CXC/GRO chemokines can exist as monomers or oligomerize into dimers/higher order oligomers [14,27]. The oligomeric state of CXCL3 was analyzed by size exclusion chromatography. The elution profile of CXCL3 was compared with standard proteins of different sizes including pepsin (36 kD), chymotrypsin (25.6 kD), cytochrome C (12 kD) and aprotinin (6 kD) and with its closest member CXCL2. It was observed that both CXCL3 and CXCL2 have been eluted at the retention time that corresponds to apparent molecular weight of ~ 16 kD when compared with retention time of standard proteins. The elution volume of CXCL2 corresponding to 16 kD is in good agreement considering its molecular weight in the homodimeric form as reported earlier [26]. SEC studies suggest that CXCL3 forms a strong homodimer like its closest paralog CXCL2 which exists in dimeric conformation **(Fig. 4.7 A)**.

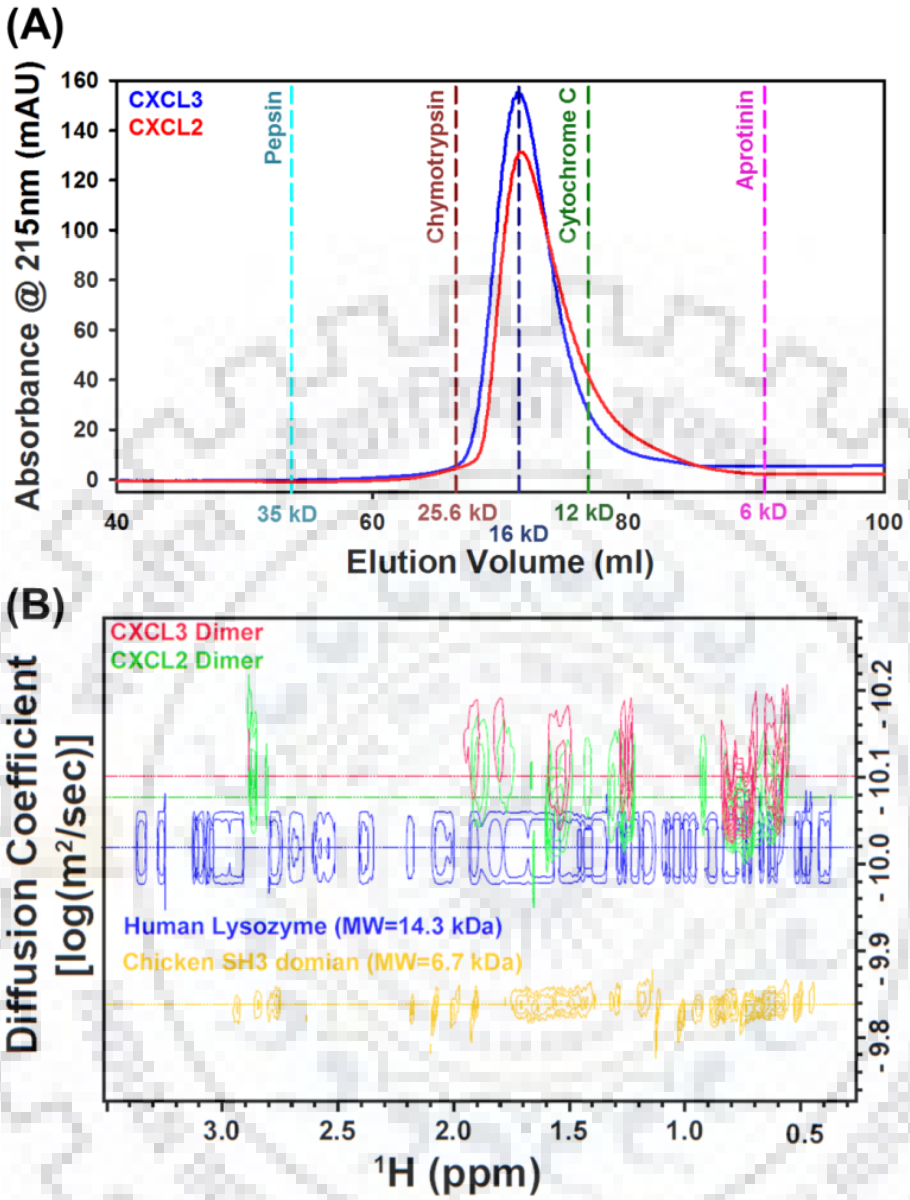


Figure 4.7: (A) *Size exclusion chromatography elution profiles of CXCL3 (blue) and CXCL2 (red). Standard marker proteins of different sizes are shown as straight lines at their elution maxima.* **(B)** *2D-DOSY analysis of CXCL2 and CXCL3 along with reference proteins hen egg lysozyme (HEL) and Chicken SH3 Domain.*

To further substantiate the oligomeric state of CXCL3, 2D-DOSY translational diffusion experiments of CXCL2 and CXCL3 were performed along with standard protein hen egg lysozyme (HEL) and chicken SH3 domain. The diffusion coefficient is related to the molecular size of the protein, which can be determined by recording series of ^{1}H NMR experiments by progressively increasing the pulse field gradient. The translation diffusion coefficient analysis of

CXCL3 and CXCL2 is shown in **Fig. 4.7 B**. It has been observed that the translational diffusion coefficients of CXCL3 and CXCL2 are higher than for chicken SH3 domain and hen egg lysozyme (HEL) indicating their slower diffusion. This slow diffusion indicates the presence of species that are more than 14kD, thus inferring the presence of dimeric form of CXCL2 and CXCL3 (~ 16kD) **(Fig. 4.7 B)**. NMR Diffusion results are in line with the SEC experiments. Further, appearance of ~ 70 resonances in the ^{1}H-^{15}N HSQC spectrum **(Fig. 4.4** and **4.6)** substantiates the homodimeric conformation of CXCL3 protein.

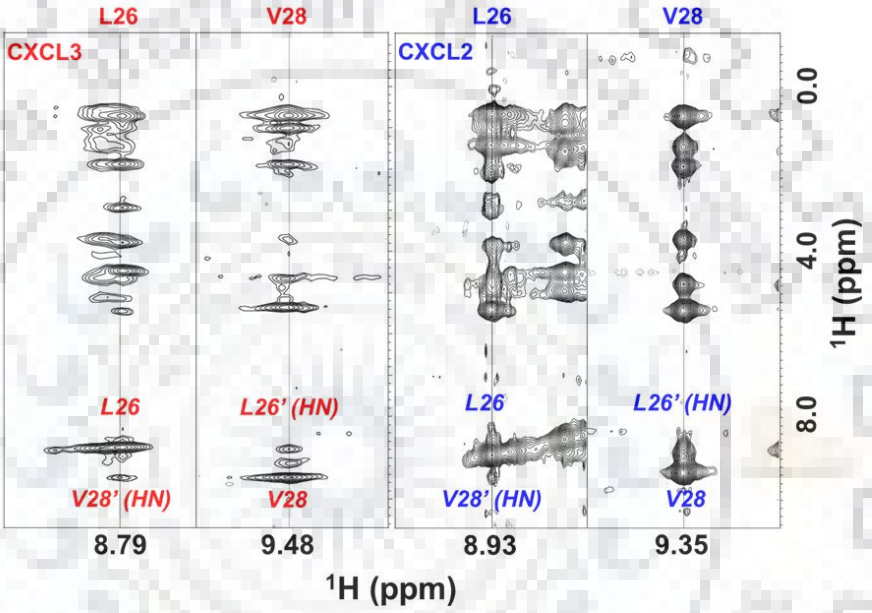


Figure 4.8: *2D- strip plots from ^{15}N- edited 3D-NOESY spectrum for the dimer interface residues L26 and V28 for CXCL3 (red) and CXCL2 (blue). The inter-strand NOE connectivities between β_1 and β_1' strands are marked.*

The geometry of the CXC chemokine dimers can be obtained using ^{15}N–edited NOESY experiment. In case of CXC dimer, one will observe the intermolecular NOE connectivities between the residues present in β_1–β_1' strand amide protons. In order to confirm the CXC type dimerization for CXCL3 chemokine, the cross peaks of the dimer interface β-strands were analyzed in a ^{15}N–NOESY experiment and were compared with the corresponding residues of the CXCL2 dimer. Two dimensional strip plot of the NOESY spectrum displaying the intermolecular NOE connectivities for the dimer interface residues L26 and V28 have been shown in **Fig. 4.8**. The similar pattern of CXCL3 NOE connectivities at dimer interface (β_1–β'_1 strand) to that of CXCL2 confirms the CXC type dimeric structure as CXCL2.

4.3.5 In-silico analysis of CXCL3 structural features

Contact map analysis was performed to analyze the differential structural contacts of CXCL3 in comparison to CXCL1 and CXCL2. Although the tertiary and quaternary structures are similar for these chemokines, the surface charges and the side chain contacts may differ due to the variation in the primary sequences **(Fig. 4.9 A)**. Contact map analysis for the aligned monomeric structure of CXCL3 yielded differential contacts when aligned with that of CXCL1 and CXCL2 **(Fig. 4.9 B-C)**. An overlap of 70 % of the contact maps was obtained for CXCL3, when analyzed using CXCL1 **(Fig. 4.9 B)**. In contrast, the contact maps of CXCL3 showed an overlap of 93.4% with CXCL2 **(Fig. 4.9 C)**. Although some contact differences were observed in comparison to CXCL1 and CXCL2, the major contacts dictate the structural fold is similar for CXCL3 in both cases. The differences in their contacts can be attributed to the differences in their local environments/amino acid sequences. Further to assess the similarity, the structure of CXCL3 was aligned with that of CXCL2. The root mean square deviation (RMSD) of 0.103 Å for the aligned structures of CXCL3 and CXCL2 indicates the overall structural similarity and higher spatial equivalence between these chemokines, as the RMSD is quantitative measurement of similarity between the Cα atomic coordinates of two superimposed protein structures **(Fig. 4.9 D)**.

This structural comparison implies the coherence between the CXCL3 and CXCL2 structures thereby, reflecting their homologous functions. However, numerous biological studies have shown that CXCL3 and CXCL2 in addition to the performance of their common function, they are involved in diversified functions and also follow differential pathways in accomplishing their common functions. This galvanized to further probe into their surface characteristics. Therefore, the electrostatic surface potential maps of CXCL2 and CXCL3 were compared. The surfaces indicated that both CXCL2 and CXCL3 are bearing mostly the positive surface at the α-helical side but they have contrasting β-sheet surfaces. CXCL3 exhibits the negative β-surface as compared to the neutral surface in CXCL2 **(Fig. 4.9 E)**. This is the resultant of presence of negatively charged residues E21, D46, E49 in CXCL3 as compared to CXCL2 which exhibits positively charged (K21, K49) or uncharged residues (G46) at these positions. In addition, the presence of altering hydrophobic residues including R58, L59, I61 in CXCL3 as compared to the L58, V59 and K61 in CXCL2 marks for their differential hydrophobic surfaces. These observations indicate that although CXCL2 and CXCL3 sharing the similar structures and

oligomerization behaviors, overall they exhibit different surfaces on both β-sheet and α-helix surfaces.

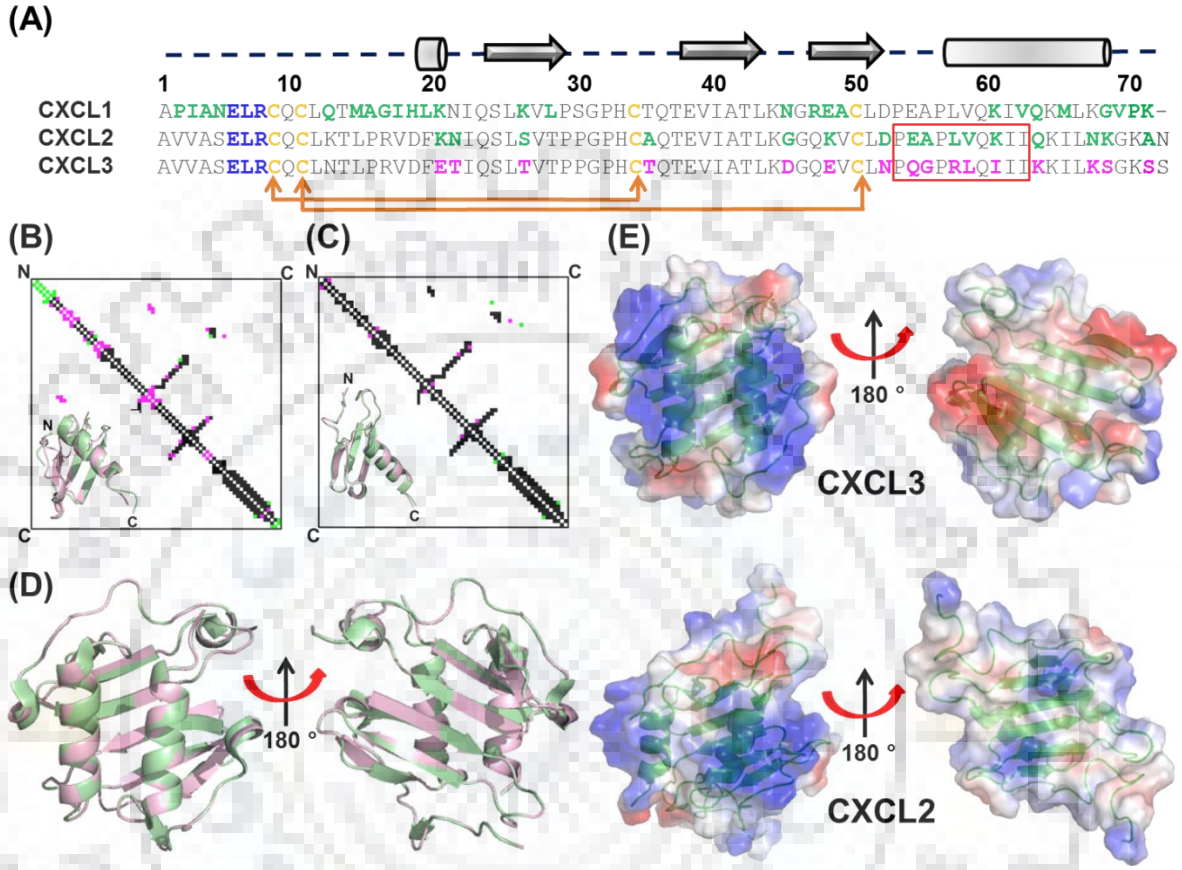


Figure 4.9: (A) *Amino acid sequence alignment of murine CXCL3 and CXCL2, in which the conserved residues are highlighted (ELR motif - Cyan, cysteine involved in disulfide bridges – Yellow). Amino acid differences are marked as pink in CXCL3 and as green in CXCL2. The hypothesized amino acid stretch at C-terminal responsible for altered hydrophobicity is enclosed using a red rectangle. The secondary structural elements are shown with arrows (β1, β2 and β3 strands) and cylinders (3_{10} - blue, α-helix). Overlay of contact maps of* **(B)** *CXCL1 (green) and CXCL3 (pink);* **(C)** *CXCL2 (green) and CXCL3 (pink); where N and C represent N-terminal and C-terminal of the protein sequences. The structural overlay of the monomeric subunits is also shown.* **(D)** *Structural alignment of CXCL2 (green) and CXCL3 (pink) generated using PyMol molecular graphic system.* **(E)** *Electrostatic surface potential maps for CXCL3 and CXCL2 in dimeric form. Both the α-helical and the β-sheet surfaces are shown. The vacuum electrostatics image was generated using PYMOL molecular graphics system.*

4.3.6 Heparin binding features of GRO chemokines

Apart from oligomerization, other important property of chemokine family is their ability to bind various glycosaminoglycans (GAGs). Chemokines bind to heparin and other glycosaminoglycans on the endothelial cell surface to regulate the leukocyte trafficking [28,29]. To estimate the efficacy of GRO chemokines for GAGs (heparin), heparin binding assay was performed by applying the proteins onto the heparin column **(Fig. 4.10)**. Distinct elution profiles were observed for GRO proteins **(Fig. 4.10 A)**. CXCL1 was eluted at a higher NaCl concentration (785 ± 20 mM) as compared to CXCL2 and CXCL3 (575 ± 15 mM) **(Fig. 4.10 B)**. These results indicate that CXCL3 and CXCL2 have a similar affinity for heparin, but weaker as compared to CXCL1.

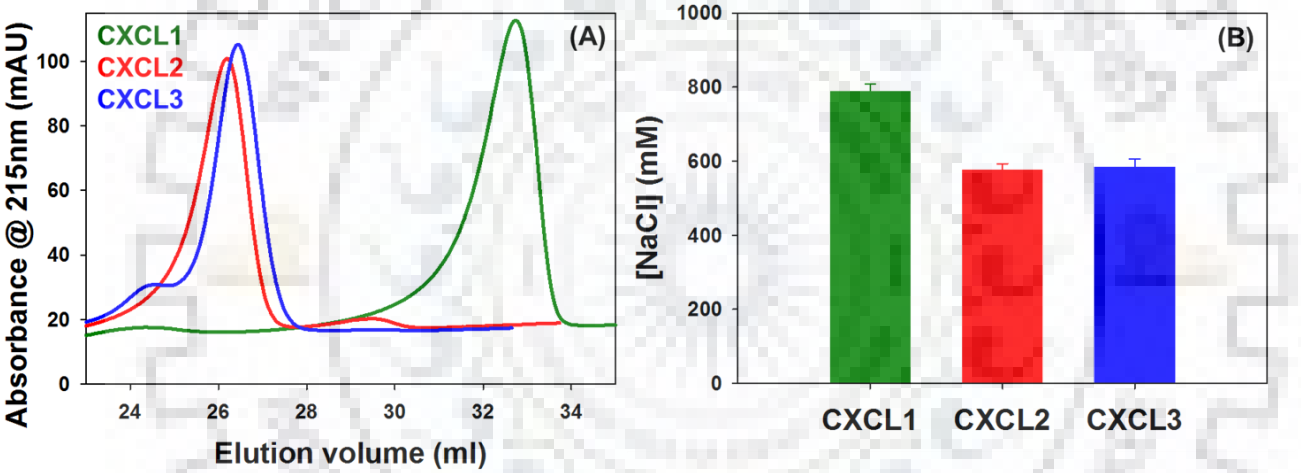


Figure 4.10: *Heparin binding assay for GRO chemokines (CXCL1, CXCL2, and CXCL3):* **(A)** *Elution profiles,* **(B)** *Bar graph representing the binding affinities.*

As described in **Chapter 3**, under the given solution conditions, CXCL1 shows two set of resonances that corresponds to the monomeric and dimeric species of the protein as compared to CXCL2 that showed only one set of resonances that corresponds to dimeric species of the protein, CXCL3 also showed only one set of homodimeric resonances similar to CXCL2. Moreover, the structural contacts of CXCL3 and CXCL2 are also highly similar in contrast to CXCL1, thus establishing the closely related structural and oligomerization features of CXCL3 and CXCL2. Considering the differential oligomeric features of CXCL1, under given conditions it is difficult to compare its biophysical features with CXCL2/CXCL3, as the population of dimer/monomer varies significantly. Therefore, in the following sections, all the biophysical characteristics of CXCL3

exclusively compared with CXCL2 to delineate the differential structural/stability features of paralog proteins.

4.3.7 Thermal stability analysis of CXCL3

Far-UV CD spectroscopy is a useful tool to investigate the secondary structural features, and also to study the effect of temperature/denaturant on the conformation of proteins [30]. CD spectra of CXCL3 and CXCL2 evidenced similar spectral features suggesting that both the proteins have presumably the same secondary structure content **(Fig. 4.11 A)**.

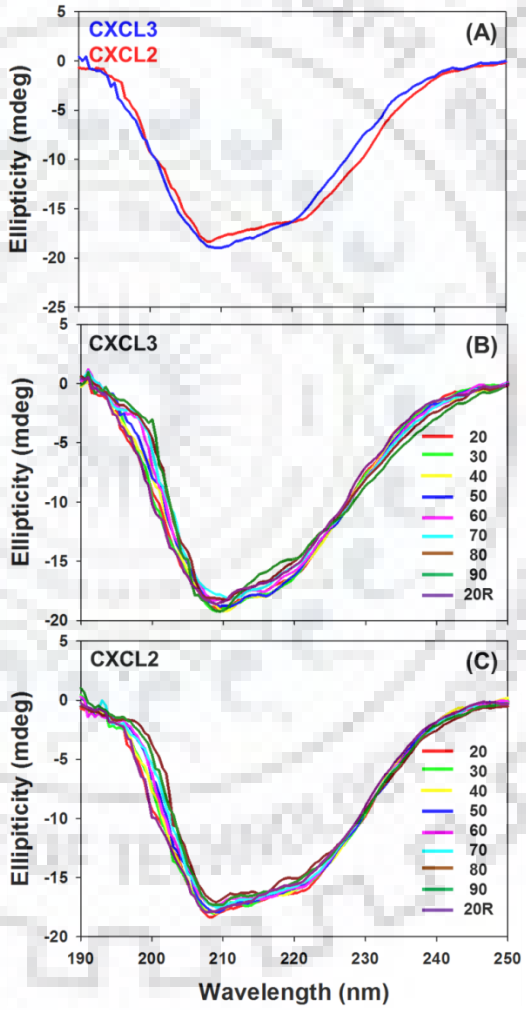


Figure 4.11: (A) *Far-UV CD profile of CXCL3 (blue) and CXCL2 (red);* **(B)** *and* **(C)** *represent the thermal denaturation CD profiles of CXCL2 and CXCL3 from 20 ºC to 90 ºC. The spectra obtained after cooling the sample back to 20 ºC is represented as 20R.*

Further, to assess their conformational stability, temperature dependent far-UV CD measurements were carried out for CXCL3 and CXCL2 **(Fig. 4.11 B-C)**. No decrease in the ellipticity was observed for both CXCL3 and CXCL2 up to 60 ºC, and ~ 10% decrease in the ellipticity (at 222 nm) was observed at around 70-90 ºC indicating that both the proteins are highly stable to thermal perturbations.

4.3.8 CXCL3 and CXCL2 binds differentially to hydrophobic probe ANS

Both the GRO chemokines CXCL2 and CXCL3 are scarce in aromatic amino acids **(Fig. 4.9 A)**. They do not possess neither Trp nor Tyr residues and contain only one Phe (F20). Hence performing a tertiary CD/intrinsic fluorescence spectroscopy measurements is out of context. Hence, in order to decipher the tertiary structural characteristics/surface properties of CXCL3 and CXCL2, an extrinsic fluorophore ANS was used. ANS fluoresces minimally in the buffer with emission maxima at 525 nm. A blue shift in the fluorescence, and an increase in the fluorescence intensity will be observed when ANS binds to the hydrophobic surface [31,32]. ANS showed distinct fluorescence profiles upon binding to CXCL2 and CXCL3 **(Fig. 4.12)**. The ANS binding to CXCL2 has resulted in the blue shift of ~ 13 nm in the spectra with 1.6 times increase in the fluorescence intensity. However, ANS binding to CXCL3 resulted in large blue shift of ~ 30 nm in the spectra with 2.7 times increase in the fluorescence intensity. Such differential spectral shifts/intensities of ANS upon binding to CXCL2 and CXCL3 indicate that their surface hydrophobicity varies considerably. Further, the results evidence that CXCL3 has more hydrophobic surface exposed as compared to CXCL2. Sequence analysis of both the chemokines with respect to the hydrophobic residues hints that the rearrangement of the polar/hydrophobic residues in the C-terminal helix of the proteins could be the reason for such an enhanced/altered ANS fluorescence characteristics of CXCL3 **(Fig. 4.9 A)**. Such differential hydrophobic surfaces can be exploited to design small molecule inhibitors for one of these two paralogs specifically. Considering the potential of rational small molecule design, we have extensively studied the ANS binding characteristics for CXCL2 and CXCL3 using time resolved fluorescence, NMR and molecular docking approaches.

Mode of binding of ANS to CXCL3 and CXCL2

To determine the binding mode of ANS to CXCL2 and CXCL3, time resolved fluorescence life time measurements were carried out for ANS alone, and in complex with CXCL2 and CXCL3.

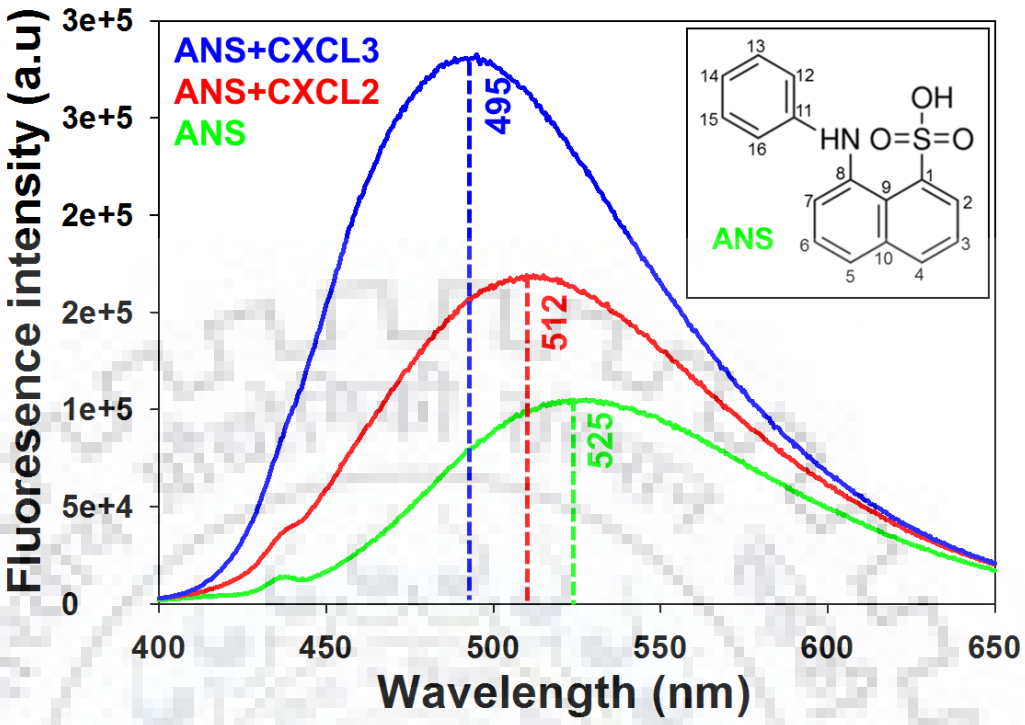


Figure 4.12: *Fluorescence spectra of ANS alone (green) and in complex with CXCL2 (red) and CXCL3 (blue). Inset shows the structure of ANS.*

Fig. 4.13 shows the life time decay profiles of ANS alone (free ANS) and for the ANS+CXCL2 and ANS+CXCL3 complex. The fluorescence life time decay of free ANS was uniexponential with a fluorescence life time of 0.26 ns. In contrast, the fluorescence life time decay of ANS in presence of CXCL2/CXCL3 was triexponential with three different time decays **(Table 4.3)**. The triexponential decay nature indicates the existence of three different species with different life times. The shortest life time decay corresponds to free ANS, whereas other two decays correspond to the ANS bound to the protein in different modes. Longer life time (τ_3) corresponds to the species in which ANS is bound to hydrophobic core and the shorter life time (τ_2) corresponds to the species in which ANS is bound to surface [33,34]. The life time decays were accompanied with different amplitudes **(Table 4.3)**. For free ANS, higher amplitude of 47 % was observed in CXCL2 as compared to the lower amplitude of 18.8 % in CXCL3. Surface bound ANS species with similar amplitudes of around 26 % was observed for both CXCL2 and CXCL3 cases. For ANS species bound to the hydrophobic core/protein embedded ANS species, amplitude of 55.1 % for CXCL3 and 26.6 % for CXCL2 was obtained. Analysis of different life time components suggest that proportion of surface bound ANS species (τ_2) was similar in both CXCL2

and CXCL3. In contrast, the proportion of protein embedded ANS species (τ_3) was much higher in case of CXCL3 as compared to CXCL2.

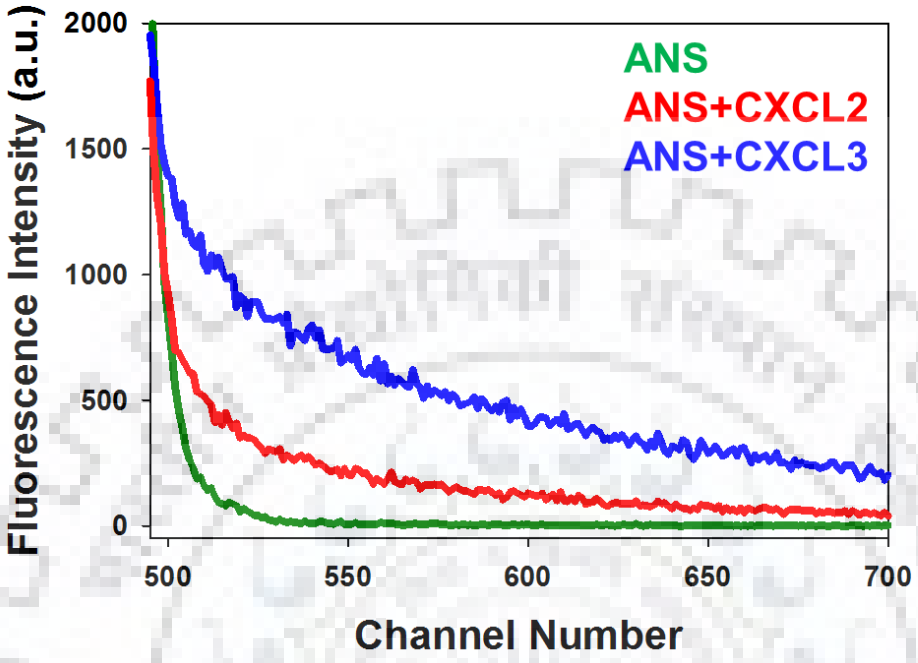


Figure 4.13: *Fluorescence lift time decays of ANS alone (green), and in complex with CXCL2 (red) and CXCL3 (blue).*

Overall, an increase in the average fluorescence life time for both ANS+CXCL2 and ANS+CXCL3 was observed in comparison to free ANS. Among two chemokines, the life time of CXCL3 is greater than CXCL2. The distinctive profile of population of different ANS species in presence of CXCL2 and CXCL3 indicates the differential interaction behavior/affinity of ANS towards CXCL2 and CXCL3.

Table 4.3: *Fluorescence life times of ANS alone (free ANS), in complex with CXCL2, and CXCL3. Relative amplitudes of each of the life time values are enclosed in parenthesis.*

Sample	**τ_1 (ns)**	**τ_2 (ns)**	**τ_3 (ns)**	**Average life time (ns)**	**χ^2**
ANS	0.26 (100)	-	-	0.26	0.99
ANS + CXCL2	0.23 (47.0)	2.4 (26.44)	10.5 (26.56)	3.53	1.04
ANS + CXCL3	0.26 (18.75)	3.33 (26.16)	11.4 (55.09)	7.25	1.07

ANS binds specifically to CXCL3 but not to CXCL2

ANS binds to proteins both specifically and non-specifically, therefore in order to further characterize the mode of ANS binding to CXCL2 and CXCL3, and to determine the residue level information of ANS binding, NMR spectroscopy was used. This chemical shift perturbation (CSP) method was used to map the binding of ANS to CXCL2 and CXCL3. No significant perturbation was observed for CXCL2, at ANS to protein molar ratio of 1:1 and 2:1. However, at the same molar ratios, significant CSP for a particular set of residues was observed in case for CXCL3 **(Fig. 4.14 and 4.15 A)**. This implies that CXCL3 has a specific binding pocket for ANS in contrast to CXCL2. **Fig. 4.15 B** showing the CSP map for CXCL3 upon binding to ANS at molar ratio of 2:1 (ANS: protein). Presentation of perturbed residues on the surface of CXCL3 clearly indicates that the ANS binds on the helical surface. No significant perturbation was noticed on β-sheet surface. Majority of the perturbed residues constitute the hydrophobic pocket (F20, I61, I62, I63) on CXCL3 helical surface **(Fig. 4.16)**.

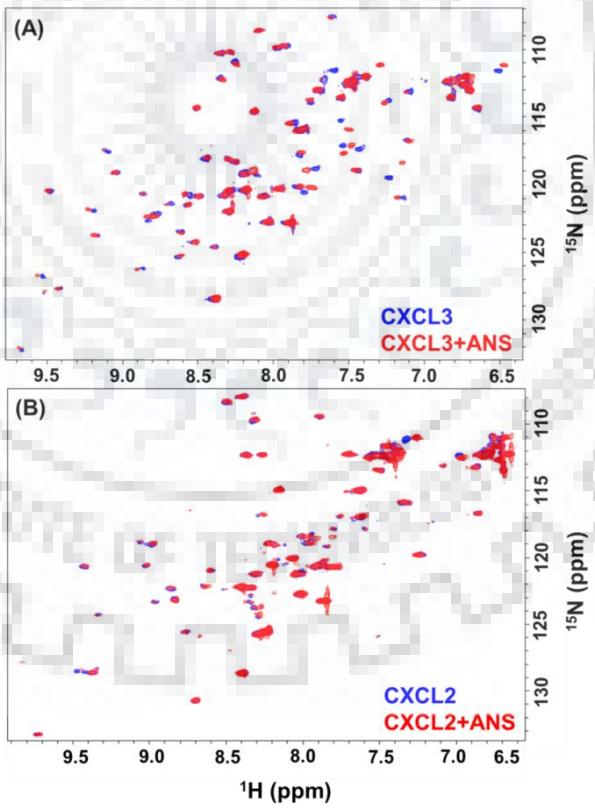


Figure 4.14: *An overlay of ^{1}H-^{15}N HSQC spectra of CXCL3* **(A)** *and CXCL2* **(B)** *alone (blue) and in presence of ANS at molar ratio of 2:1 (ANS:Protein) (red).*

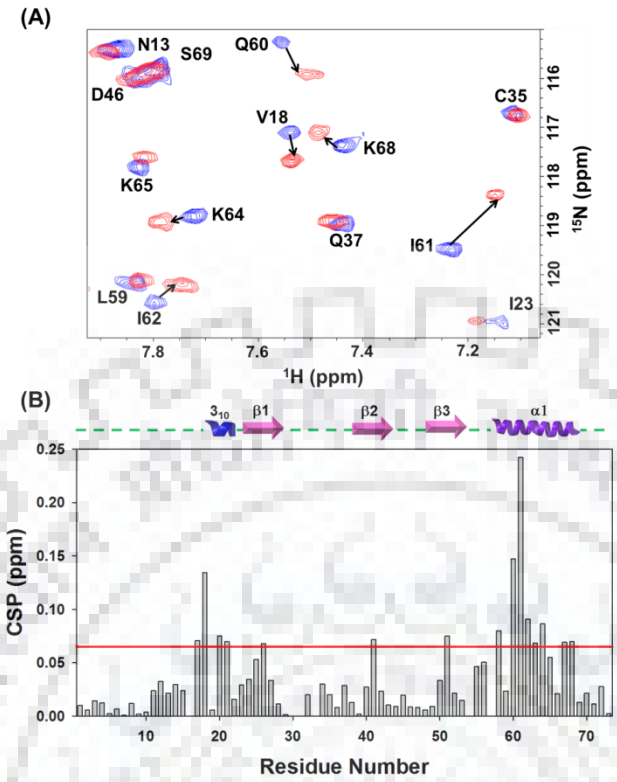


Figure 4.15: **(A)** *A section of ^{1}H-^{15}N HSQC spectrum of free CXCL3 (blue) and in the presence of ANS (red) with arrows indicating the direction of change in chemical shift induced by ANS binding.* **(B)** *Chemical shift perturbation (CSP) map of CXCL3-ANS interactions. The horizontal red line at 0.07 indicates the cut off for the residues to be considered as perturbed. The secondary structural elements have been shown on the top of the chemical shift map with arrows for β-sheets and helix for 3$_{10}$ and α-helix.*

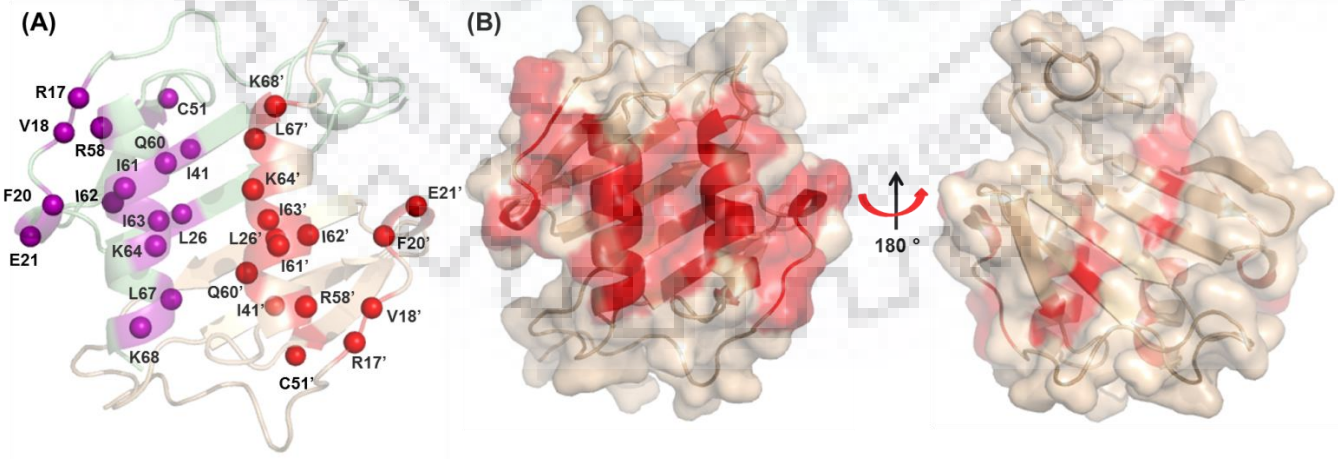


Figure 4.16: *CXCL3 residues that are significantly perturbed on ANS binding are shown as spheres* **(A)** *and marked as red on the surface of CXCL3;* **(B)** *Both α-helical and β-sheet surfaces have been shown.*

ANS:CXCL3 binding stoichiometry

In order to determine the binding stoichiometry of ANS to CXCL3, Job plot (continuous variation plot) analysis was carried out. In this method, the mole fraction at point of inflection yields the binding stoichiometry between the interacting molecules as described in **Section 4.2.6**. The maxima of the curve or the point of inflexion in the plot lying at the mole fraction of 0.5 indicates the binding stoichiometry of 1:1, where the CXCL3 concentration is represented as monomer **(Fig. 4.17)**.

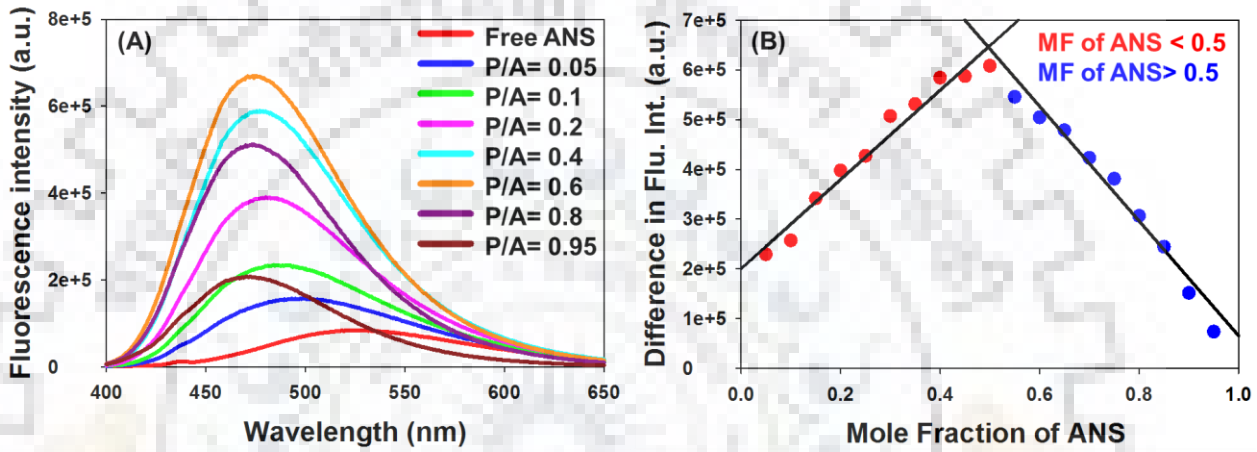


Figure 4.17: (A) *Fluorescence emission spectra of ANS alone and in presence of CXCL3 at different protein: ANS (P/A) mole fractions,* **(B)** *Job plot for binding of ANS to CXCL3 using ANS fluorescence intensity for a fixed total concentration (100 µM) of ANS and CXCL3.*

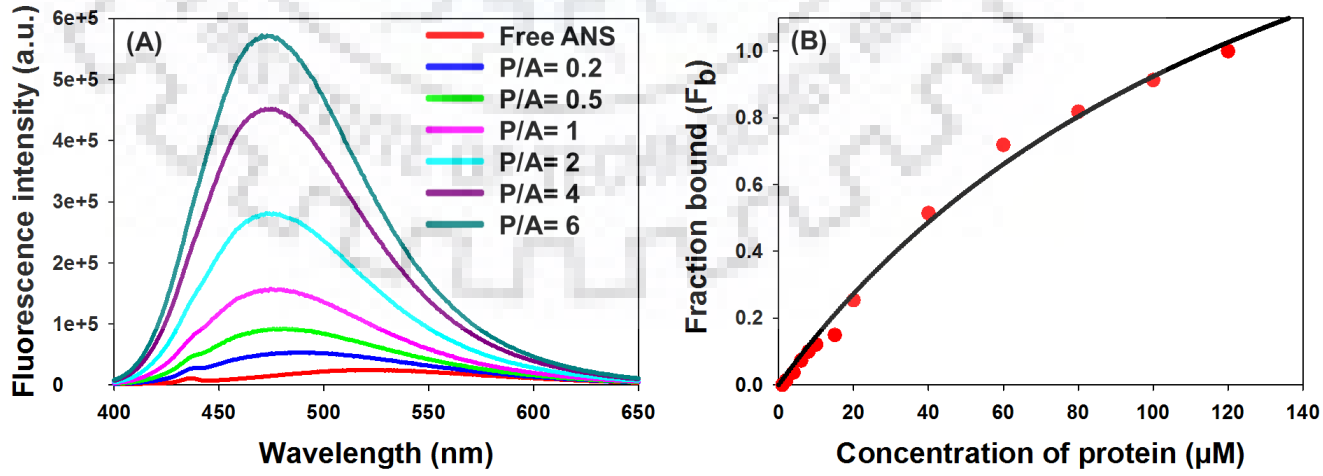


Figure 4.18: (A) *Fluorescence spectra of free ANS (100 µM) and with increasing concentration of CXCL3 with protein:ANS (P:A) molar concentration ranges from 0.2 to 6,* **(B)** *Increase in*

fluorescence intensity was monitored for ANS-CXCL3 binding with increasing CXCL3 concentration.

This implies that there is one ANS binding to one CXCL3 monomer. To further determine the binding constant for ANS:CXCL3 complex, ANS fluorescence titration measurements were carried out. For the binding constant, ANS (100 µm) was titrated with CXCL3 (1 to 100 µM) and fitted to a hyperbolic curve. A dissociation constant (K_d) of 97±14 µM was obtained for the binding interaction **(Fig. 4.18**). The NMR and fluorescence data suggest that although ANS binds in the specific pocket of CXCL3, the binding is inherently weak **(Fig. 4.18**).

Molecular docking of ANS to CXCL3

The molecular docking has been carried out considering the 1:1 stoichiometry of ANS:CXCL3 monomer by presuming that the binding site of ANS molecule to monomeric units of CXCL3 dimer are independent and identical. The consideration is based on the NMR data which indicated the binding site is far away from the CXCL3 dimer interface. Docking has been performed by using CXCL3 monomeric structure. Docking results demonstrated that the ANS binds into the groove formed by the various residues present on the C-terminal helix and 3_{10} helix with the binding energy of -5.8 kcal/mol **(Fig. 4.19 A and B)**. In general, ANS with its sulfonated naphthalene group and aniline group interacts with proteins through various interactions including: (a) Ionic interactions: Sulfonate group of ANS interacts electrostatically with positively charged amino acids, (b) Hydrophobic interactions: the aromatic naphthalene and aniline rings interacts hydrophobically with non-polar amino acids, (c) Hydrogen bonding: the amide group of aniline ring can form hydrogen bonds with protein [35]. In the present case, ANS is locked into CXCL3 as a result of π-π interactions of phenylalanine (F20) benzene ring, and CH-π interactions of I61 and I62 methyl groups with naphthalene ring of ANS. ANS is also stabilized by the electrostatic interaction of sulfonate group of ANS with guanidium group of R58. Additional hydrophobic interactions of ANS with K64, K65, and E21 played an assistive role in binding of ANS with CXCL3. All the possible interactions of ANS to CXCL3 have been summarized in **Table 4.4** and some of the essential contacts have been depicted on the structure (**Fig. 4.19 C)**. The binding geometry of ANS onto CXCL3 is consistent with the NMR chemical shift perturbation experiments. Among the significantly perturbed residues in CSP plot of CXCL3:ANS (**Fig. 4.15 B)**, R17 and V18 did not yielded any direct interaction with ANS. The distance measurements suggested that side chains of V18 and I62 are in close contact (3.4 Å). Thus the observed changes

may be due to (a) secondary/indirect chemical shift effects, (b) nonspecific binding/secondary binding site of ANS in this region. However, no such significant shifts were seen for R17 and V18 residues of CXCL2 indicating that these residues were perturbed due to secondary effect/conformational adjustments. Thus, the combined fluorescence, NMR, and docking studies establish that ANS binds specifically to CXCL3 monomer in hydrophobic groove formed by residues of both C-terminal and 3_{10} helix with a stoichiometry of 1:1.

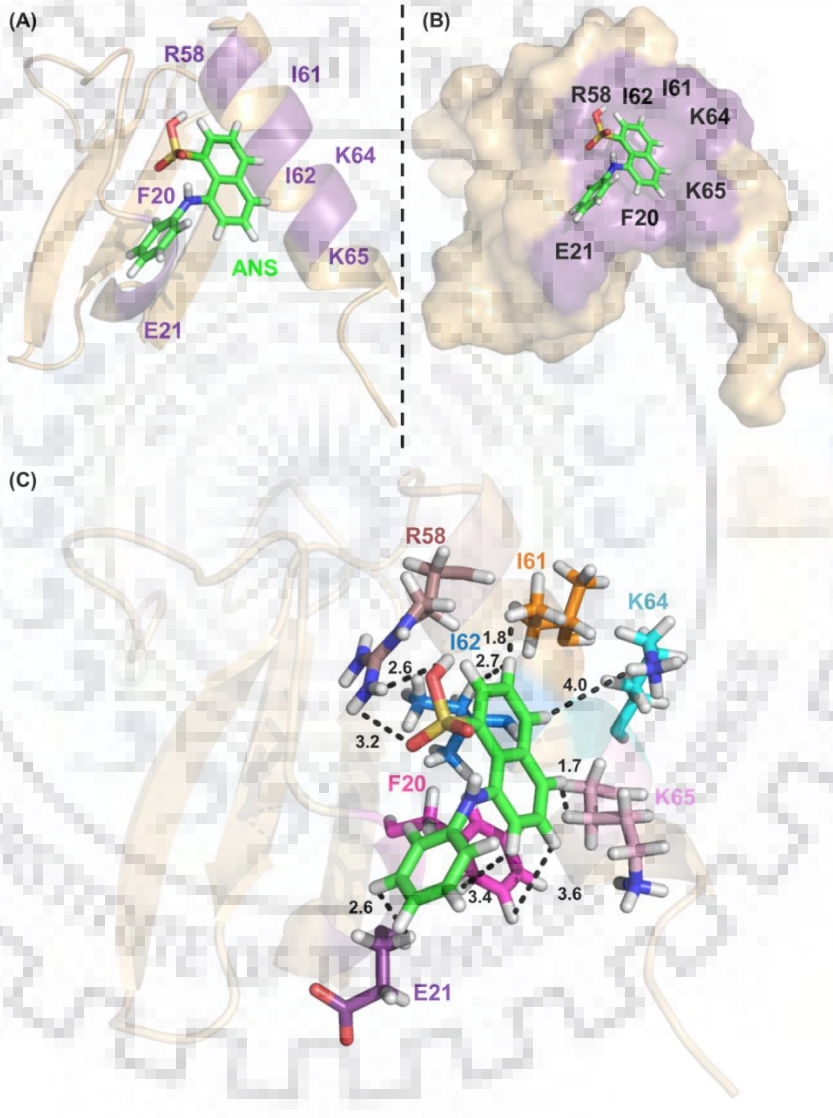


Figure 4.19: (A) *Docking of ANS (green) into the modeled CXCL3 monomer structure (light orange) using AutoDock Vina. Residues involved in interaction with ANS are marked on the structure in purple,* **(B)** *CXCL3 surface structure representing the residues forming a hydrophobic pocket for interacting with ANS,* **(C)** *CXCL3-ANS complex structure depicting various interactions between the side chains of CXCL3 residues with naphthalene, sulphonic acid, and aniline rings of ANS as represented by dotted lines. The number indicates the distance between the atoms in Å.*

Table 4.4: *Summary of various interactions observed between CXCL3 and ANS.*

CXCL3	ANS	Distance (Å)	Type of interaction
R58 (η1NH1)	S (OH)	2.6	Electrostatic
R58 (η1NH2 or 1')	S (O)	3.2	Electrostatic
R58 (δCH)	C2H	2.4	Hydrophobic
R58 (γCH)	C3H	2.7	Hydrophobic
R58 (αCH)	C4H	2.7	Hydrophobic
R58 (βCH)	C4H	3.9	Hydrophobic
R58 (βCH)	C4H	4.8	Hydrophobic
R58 (αCH)	C5H	5.8	Hydrophobic
I61 (δCH1)	C3H	1.8	Hydrophobic
I61 (δCH2)	C4H	3.7	Hydrophobic
I61 (βCH2)	C5H	6.0	Hydrophobic
I62 (δCH1)	C3H	2.7	Hydrophobic
I62 (δCH1)	C4H	3.0	Hydrophobic
I62 (δCH1)	C2H	3.9	Hydrophobic
I62 (γ1CH1)	C2H	4.4	Hydrophobic
I62 (γ2CH1)	C5H	5.2	Hydrophobic
K64 (ζNH1)	C4H	4.0	Van der Waal
K64 (εCH)	C3H	5.8	Hydrophobic
K65 (δCH)	C5H	1.7	Hydrophobic
K65 (εCH)	C6H	2.2	Hydrophobic
K65 (ζNH)	C7H	5.7	Van der Waal
F20 (ε1CH)	C6H	3.6	Hydrophobic
F20 (ε2CH)	C5H	3.9	Hydrophobic
F20 (δ1CH)	C5H	3.4	Hydrophobic
F20 (δ1CH)	C16H	3.7	Hydrophobic
E21 (βCH)	C15H	2.6	Hydrophobic
E21 (γCH)	C14 H	3.6	Hydrophobic

All the above analysis on surface characteristics clearly indicated that the paralogs CXCL2 and CXCL3 show significant variations due to varied electrostatic potentials. Such variation in the electrostatic potentials not only influences the surface properties, but also can modulate their dynamics and stabilities, as several of these residues are involved in tertiary/quaternary interactions. Hence, in order to unravel the residue level differences in dynamics, stabilities and conformational fluctuations of these two paralogs, NMR based relaxation, hydrogen exchange and temperature dependent amide proton chemical shift studies were performed.

4.3.9 Conformational dynamics of CXCL3 and CXCL2

Backbone NMR relaxation measurements provide information on residue specific motions of a protein at different time scales. In more generic terms, longitudinal relaxation rate ($R_1 = 1/T_1$), is sensitive to both low and high frequency motions but it alone cannot effectively discriminate

between the faster and the slower motions. Transverse relaxation rate ($R_2 = 1/T_2$) is sensitive for slower motions ranging from millisecond to microsecond time scales. Steady state heteronuclear

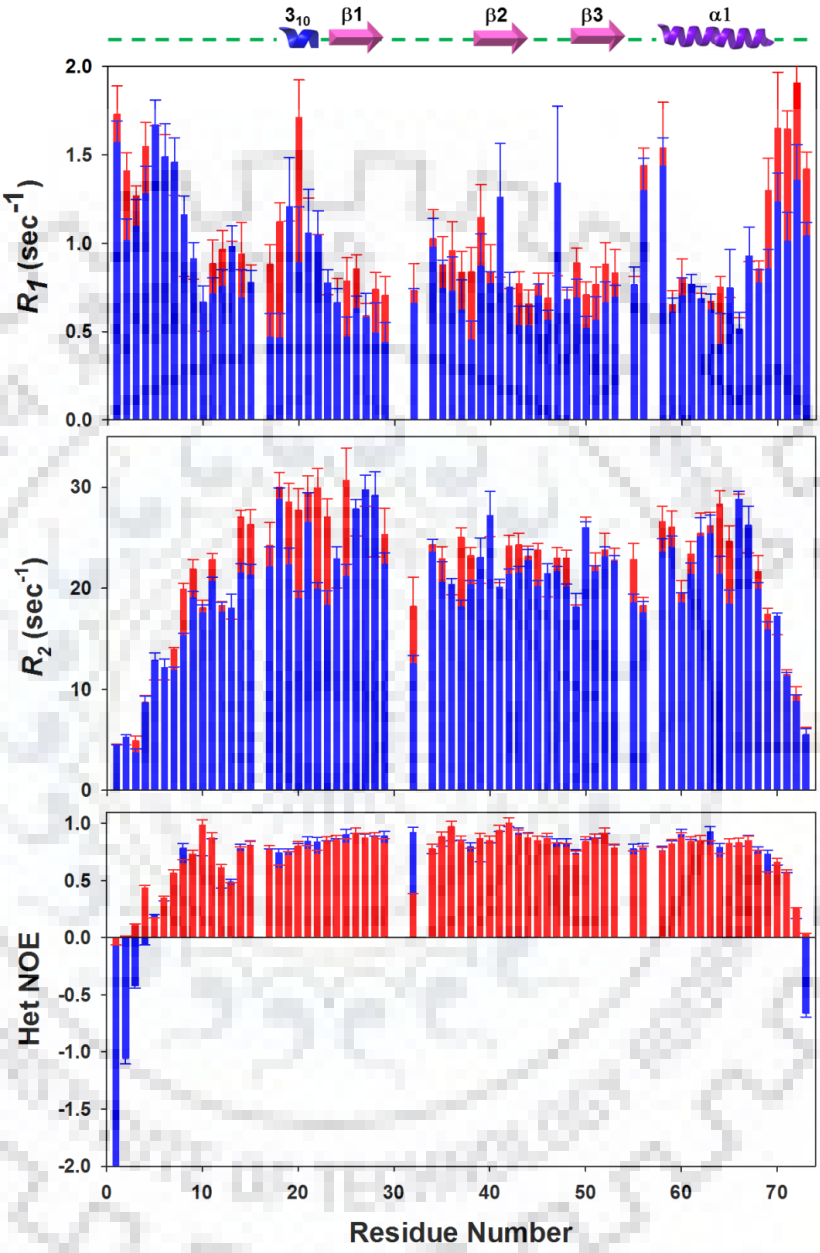


Figure 4.20: *Overlay of* **(A)** *longitudinal relaxation rates (R_1),* **(B)** *Transverse relaxation rates (R_2), and* **(C)** *steady state heteronuclear NOE (Het-NOE) of residue wise ^{15}N relaxation parameters of CXCL3 (red) and CXCL2 (blue). The secondary structure elements in the protein are marked as arrows for β-sheet and helices for 3_{10} and C-terminal α-helix.*

NOEs (Het NOE) are sensitive to high frequency motions i.e. picosecond to nanosecond time scale motions. Its lower value indicates higher local flexibility in the polypeptide backbone. Backbone

^{15}N relaxation experiments were performed to unravel the conformational dynamics of CXCL2 and CXCL3. Relaxation parameters (R_1, R_2, NOE) obtained for each of the residues in both CXCL2 and CXCL3 have been shown in **Fig 4.20**. Both the longitudinal and transverse relaxation rates showed systematic variations along the polypeptide chain (**Fig. 4.20 A-B**). The average R_2 values for CXCL2 and CXCL3 are 21.1 ± 1.2 and 19.6 ± 1.0 s^{-1}. CXCL3 exhibits the comparable R_2 values with CXCL2 for the structural elements β_2, β_3, and differential dynamic features at the N-loop, 3_{10} helix, β_1 strand, and C-terminal helix. The enhanced R_2 values of these structural elements in CXCL3 depict their ms-μs flexibility, thus suggesting to have considerable contribution from the conformational exchange phenomenon to the R_2 values. A set of CPMG dispersion experiments are essential to visualize the location and extent of exchange phenomenon in these residues.

Furthermore, heteronuclear NOEs with an average value (residues 5-72) of 0.74 ± 0.05 and 0.75 ± 0.04 for CXCL3 and CXCL2 respectively indicates the overall folded structure of the proteins (**Fig. 4.20 C**). No significant differences in the ps-ns time scale motions were observed in any of the structural elements of CXCL2 and CXCL3. The significant differences are seen only on N- and C- terminal residues. The terminal residues of CXCL3 are more flexible than CXCL2 and attained large negative NOE values.

4.3.10 Residue wise conformational stabilities of CXCL3 and CXCL2 using hydrogen exchange NMR

Native state hydrogen/deuterium exchange (NHX/HX) NMR was used to study the residue wise conformational stabilities of CXCL2 and CXCL3. In general, the residues accessible to the solvent, exchange their amide protons with solvent deuterons faster than the protons that are either present in the hydrophobic core or hydrogen bonded. NMR based native state hydrogen exchange (HX) technique can be used to determine the solvent accessibility of every amide bond present in the protein, thus providing the information regarding their stabilization free energies [36]. NHX studies were carried out for CXCL3 and CXCL2 by recording series of NMR spectra at regular time intervals for about 40 hours. **Fig. 4.21 A** and **C** show the HSQC spectra for CXCL3 and CXCL2 before adding D_2O. **Fig. 4.21 B** and **D** show the spectra of CXCL3 and CXCL2 recorded 6 minutes after adding D_2O. The spectra showed 17 peaks for CXCL3 and 20 peaks for CXCL2. All the remaining peaks have disappeared in spectra for both CXCL2 and CXCL3 within the dead time

of 6 minutes, which indicate that these residues are highly accessible to the solvent. For CXCL3, out of 17 protected residues, 10 were residing in beta strands (β_1, β_2, and β_3), 6 in interconnecting loops, and 1 in the C-terminal helix. For CXCL2, Out of 20 protected residues, 17 residues that were protected for CXCL3 were also protected in CXCL2. The additional three residues belong to 40S loop (G46), and C-terminal helix (I63, I66).

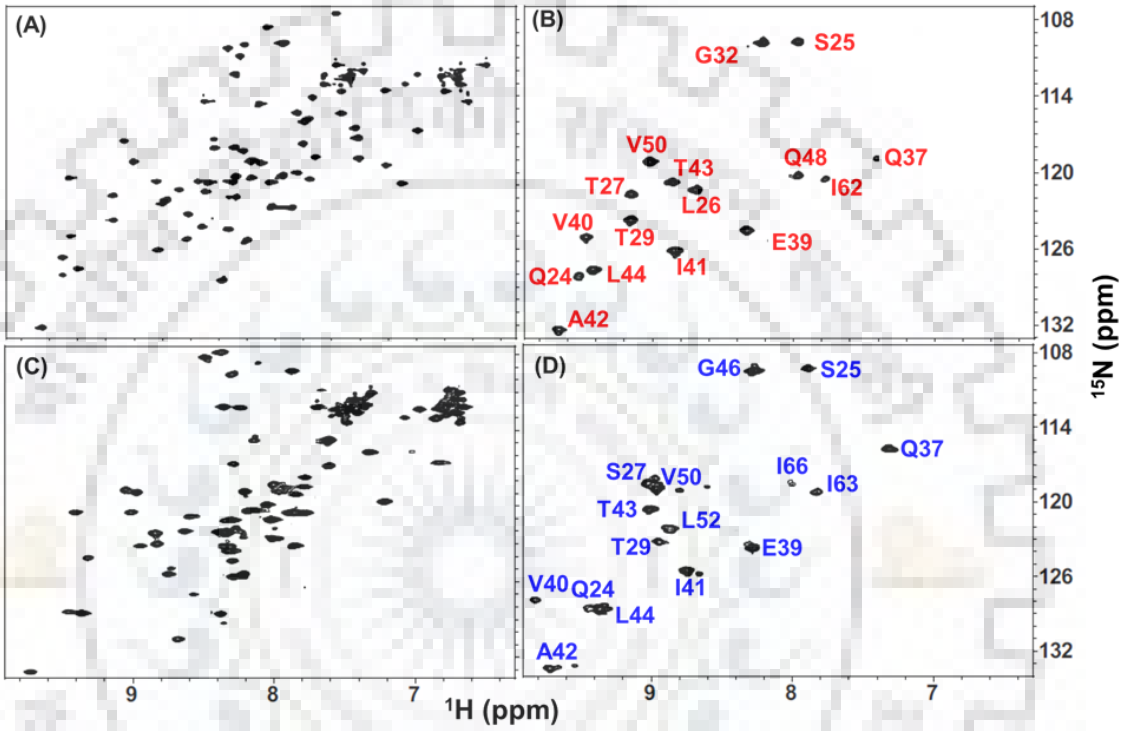


Figure 4.21: *^{1}H-^{15}N HSQC spectra of CXCL3* **(A, B)** *and CXCL2* **(C, D)** *with* **(A)** *and* **(C)** *were recorded before adding D_2O and* **(B), (D)** *were recorded 6 minutes after adding D_2O. Peaks for CXCL3 (B) and CXCL2 (D) have been marked in red and blue respectively.*

The intensity decay profiles of the protected residues for CXCL3 and CXCL2 were monitored by following the decay kinetics. After 40 hours, no peaks were left with observable intensity in CXCL3. However, around 8-10 peaks in CXCL2 have not at shown any significant intensity change even after 40 hours of exchange process. The intensity decay of the observed peaks was fitted into the single exponential decay to determine their exchange rates (k_{ex}) **(Table 4.4)**. Depending on k_{ex} values the residues have been divided into three different classes: (1) fast exchanging residues ($k_{ex} > 10^{-2}$) (2) intermediate exchanging residues ($10^{-4} < k_{ex} < 10^{-2}$), and (3) slow exchanging residues ($k_{ex} < 10^{-4}$). The residues falling in each of these categories for both CXCL3 and CXCL2 have been summarized in **Table 4.4**. In CXCL3, no residue was observed

under slowly exchange but 5 residues were undergoing slow exchange in CXCL2. In contrast, majority of the residues that are under fast exchange in CXCL3 have been observed in the intermediate exchange category in case of CXCL2. For demonstration, two such residues (A42 and Q24) with altered exchange rates in CXCL2 and CXCL3 are shown in **(Fig. 4.22)**. These results indicate that the overall exchange phenomenon in CXCL3 is faster compared to its paralog CXCL2. Residues belonging to each k_{ex} category have been marked with different colors on the surface of CXCL2 and CXCL3 **(Fig. 4.23)**. The results indicated that majority of the protected residues are lying in the β-surface in both CXCL2 and CXCL3.

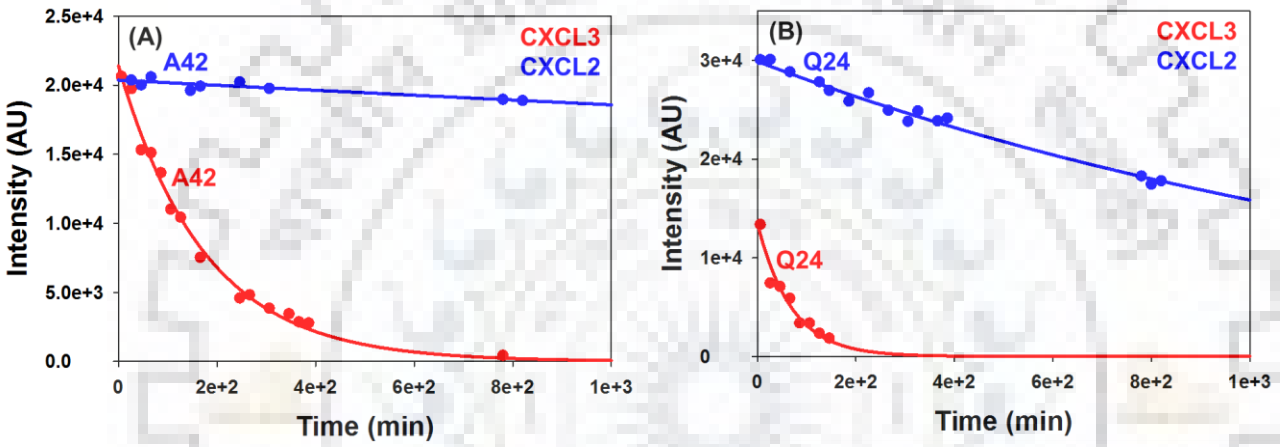


Figure 4.22: *Differential decay profiles of residues in CXCL3 and CXCL2:* **(A)** *A42 and* **(B)** *Q24.*

Table 4.5: *Summary of exchange rate patterns for various residues in CXCL3 and CXCL2*

Rate exchange (k_{ex})(min^{-1})	CXCL3 Residues	CXCL2 Residues
$k_{ex} < 10^{-4}$ (Slow exchanging residues)	-	L26, I41, A42, V50, L52
$10^{-4} < k_{ex} < 10^{-2}$ (Intermediate exchanging residues)	T29, E39, V40, I41, A42, L44, L52, L26, T43, V50	Q24, S25, T29, S27, E39, Q37, V40, T43, G46, Q48, I63, L44,
$k_{ex} > 10^{-2}$ (Fast exchanging residues)	Q24, S25, G32, Q37,T27, Q48, I62	I62, I66, G32

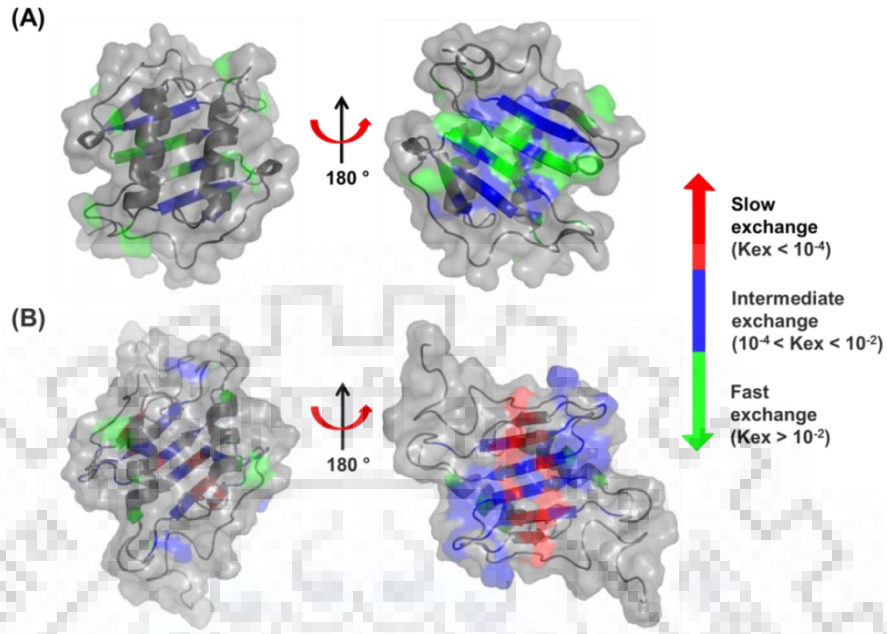


Figure 4.23: *Summary of exchange rates of individual residues marked on the surface of CXCL3* **(A)** *and CXCL2* **(B)**. *The residues marked in differential colors indicate the different range of exchange rates (k_{ex}). The color codes are as follows; Red: slow exchange rate (k_{ex} <10^{-4}), Blue: intermediate exchange rate 10^{-4} < k_{ex} <10^{-2}), and Green: fast exchange rate (k_{ex} > 10^{-2}).*

Table 4.6: *Summary of parameters calculated for individual residues in CXCL3 using HX.*

Position	Res	k_{obs} (min^{-1})	k_{rc} (min^{-1})	LogP	K_{op}	ΔG_{HX} (Kcal/mol)
N-Loop	Q24	1.4e-2 ±1.7e-3	1.1	2.9±0.3	1.3e-3	3.9± 0.1
β_1	S25	2.3e-2±1.5e-2	6.1	3.4 ±0.1	3.7e-4	4.7 ±0.1
β_1	L26	6.2e-3±1.9e-4	8.7	3.1 ±0.1	7.2e-4	4.3 ±0.1
β_1	T27	1.0e-2±5.3e-4	8.7	2.9 ±0.1	1.2e-3	4.0 ±0.2
β_1	T29	5.4e-3±2.1e-4	1.0	3.3 ±0.2	5.3e-4	4.5 ±0.2
30S Loop	G32	1.1e-2±0.8683	1.7	3.2 ±0.1	6.3e-4	4.4 ±0.1
30S Loop	Q37	0.0147±5.2e-3	3.0	3.3 ±0.1	4.9e-4	4.5±0.1
β_2	E39	5.1e-3±2.6e-4	1.1	3.3 ±0.1	4.8e-4	4.5 ±0.2
β_2	V40	7.1e-3±3.8e-4	2.7	2.6 ±0.1	2.6e-3	3.5 ±0.1
β_2	I41	1.1e-3±4.4e-5	2.2	3.3 ±0.1	4.9e-4	4.5 ±0.1
β_2	A42	5.7e-3±2.5e-4	9.6	3.2 ±0.4	5.8e-4	4.4 ±0.1
β_2	T43	7.1e-4±2.6e-5	1.4	4.3 ±0.2	5.1e-5	5.9± 0.3
40S Loop	L44	4.7e-3±2.2e-4	6.9	3.2 ±0.1	6.8e-4	4.3± 0.1
40S Loop	Q48	3.1e-2±5.5e-3	2.8	2.9 ±0.1	1.1e-3	4.0± 0.1
β_3	V50	7.1e-4±2.5e-5	2.7	3.6 ±0.1	2.6e-4	4.8± 0.1
50S Loop	L52	6.2e-3±1.9e-4	1.5	3.4 ±0.1	4.5e-4	4.6 ±0.1
α-helix	I62	2.6e-2±5.6e-3	1.8	1.8±0.1	1.5e-2	2.5±0.1

Table 4.7: *Summary of parameters calculated for individual residues in CXCL2 using HX.*

Position	Res	k_{obs} (min^{-1})	k_{rc} (min^{-1})	LogP	K_{op}	ΔG_{HX} (Kcal/mol)
N-Loop	Q24	6.3e-4±3.8e-5	1.1	4.2±0.3	5.7e-5	5.8±0.1
β_1	S25	7.5e-3±1.4e-3	6.1	3.9±0.1	1.2e-4	5.3±0.1
β_1	L26	2.8e-5±1.4e-6	8.7	5.5±0.1	3.3e-6	7.5±0.1
β_1	S27	2.4e-4±1.9e-5	2.4	5.0±0.1	9.9e-6	6.8±0.2
β_1	T29	2.3e-4±3.9e-5	1.0	4.6±0.2	2.3e-5	6.3±0.2
30S Loop	G32	9.5e-3±1.5e-4	1.7	3.3±0.1	5.5e-4	4.5±0.1
30S Loop	Q37	3.6e-3±1.7e-4	1.9	3.7±0.1	1.9e-4	5.1±0.1
β_2	E39	4.1e-3±1.3e-4	1.1	3.4±0.1	3.8e-4	4.7±0.1
β_2	V40	2.4e-3±2.6e-4	2.7	3.1±0.1	8.9e-4	4.2±0.2
β_2	I41	8.3e-6±8.4e-6	2.2	5.4±0.1	3.7e-6	7.4±0.1
β_2	A42	9.2e-5±1.2e-5	9.6	5.0±0.1	9.5e-6	6.8±0.1
β_2	T43	3.5e-3±2.1e-5	1.4	3.6±0.4	2.5e-4	4.9±0.3
40S Loop	L44	6.4e-4±3.9e-5	6.9	4.0±0.2	9.2e-5	5.5±0.1
40S Loop	G46	4.2e-3±1.5e-4	4.0	3.9±0.1	1.1e-4	5.4±0.1
40S Loop	Q48	2.9e-3±6.6e-4	2.8	3.9±0.1	1.0e-4	5.4±0.1
β_3	V50	5.5e-05±6.6e-06	4.3	4.9±0.1	1.3e-5	6.7±0.1
50S Loop	L52	2.9e-05±1.5e-06	1.5	5.7±0.1	1.8e-6	7.8±0.1
α-helix	I62	4.0e-02±2.0e-03	4.0	2.0±0.1	1.0e-2	2.7±0.1
α-helix	I63	4.9e-03±7.5e-04	1.8	2.6±0.1	2.7e-3	3.5±0.1
α-helix	I66	1.1e-02±4.1e-03	4.0	2.6±0.1	2.7e-3	3.5±0.1

Based on the exchange rates of different residues in CXCL3 and CXCL2, parameters such as protection factors and stabilization free energies (ΔG_{HX}) were calculated using the equations described in **Section 4.3.7**. The thermodynamic parameters for each of the decaying residues in CXCL3 and CXCL2 have been summarized in **Tables 4.5** and **4.6** respectively. The obtained protection factors and stabilization free energies for all the calculated residues CXCL3 and CXCL2 were shown in **Fig. 4.24** and **Fig. 4.25**. Higher protection factor and stabilization free energies have been observed for the residues lying in the β-sheets as compared to the residues in the loop and C-terminal helix. Among the β-sheets, residues lying in the β_2 strand were much more protected as compared to β_1 and β_3. Indeed the β_2 strand is sandwiched between the β_1 and β_3 strands, and its residues are tightly H-bonded. Interestingly, lower protection factors and stabilization free energies were observed for CXCL3 as compared to CXCL2 indicating that the paralogs have differential stability characteristics.

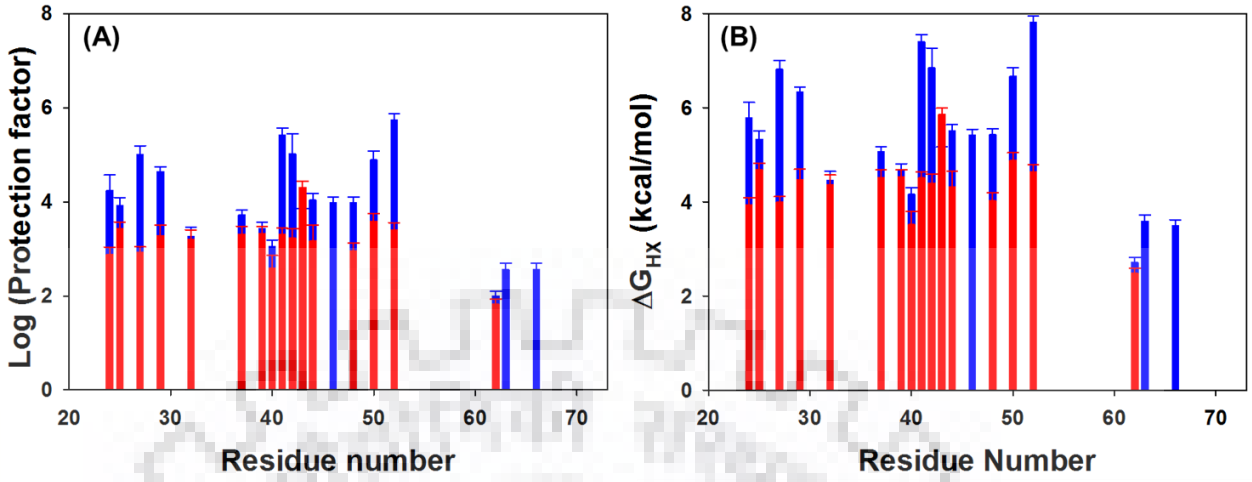


Figure 4.24: *Summary of* **(A)** *protection factors and* **(B)** *stabilization free energies for the individual residues of CXCL3 (red) and CXCL2 (blue).*

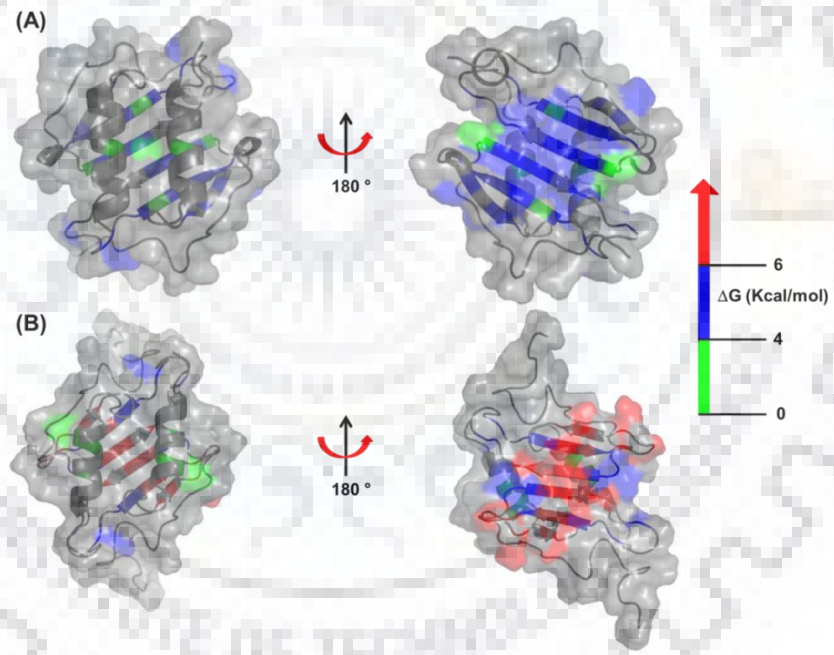


Figure 4.25: *Summary of stabilization free energies of individual residues marked on the surface of CXCL3* **(A)** *and CXCL2* **(B)**. *The residues marked in differential colors indicate the different range of stabilization free energies (k_{ex}). The color codes are as follows; Red: ΔG > 6 Kcal/mol, Blue: 6 Kcal/mol< ΔG > 4 Kcal/mol, and Green: ΔG < 4 Kcal/mol.*

The higher stability of the CXCL2 is also evident from the number of protected residues in the C-terminal helix. As shown in **Table 4.6** and **4.7**, three residues (I62, I63, I66) from helix were protected in CXCL2 in contrast to only one protected residue (I62) of CXCL3. The protection of

extra residues in the C-terminal helix and the enhanced stabilization free energies in β-sheet of CXCL2 suggests that its quaternary packing is more optimal compared to its paralog CXCL3.

4.3.11 Exploring alternative conformational states of CXCL3 and CXCL2

Proteins, in addition to their ground state conformations, can also access some near native states (within ΔG = *2-3 Kcal*/mol) that can be defined as alternative states or low energy excited states [37,38]. Considering the above differential dynamics and stabilities of CXCL2 and CXCL3, it is anticipated that they access variable alternative conformations. In order to probe the nature of the alternate conformation pattern of these two paralogs, temperature dependence NMR experiments were performed.

The Temperature dependence of the amide proton chemical shifts in CXCL3 and CXCL2 were measured by recording seven HSQC spectra as a function of temperature in the range 285-320K. Both the proteins were quite stable in the measured temperature range as indicated by the thermal CD studies in **Section 4.3.7**. This implies that the temperature dependences of CXCL3 and CXCL2 can indeed determine their intrinsic stabilities and local perturbations. The temperature dependence of some of the residues from different parts of CXCL2 and CXCL3 are shown in **Fig. 4.26 A-C**.

Variation in the chemical shift gradients were observed between these two proteins. For example, residues T27 (CXCL3), S27 (CXCL2) showed positive temperature dependencies and I41 in both the proteins showed negative temperature dependencies, although all the residues are present in the β-sheets (**Fig. 4.26 A-B)**. In contrast, to the above pattern, residues K68 (CXCL3) and N68 (CXCL2) of C-terminal helix showed opposite patterns of temperature dependencies **(Fig. 4.26 C)**. In order to compare the residue level chemical shift gradients of CXCL2 and CXCL3, the temperature coefficients of all the residues have been analyzed and plotted **(Fig. 4.26 D)**. The plot suggests that overall CXCL2 and CXCL3 exhibit similar temperature dependence profiles. However, some of the residues in CXCL3 including V18, T36, L26, K45, L52, N53, R58, K71, S72 showed comparatively more negative temperature dependence than CXCL2. The temperature coefficients of all these residues are lying in the range of -5 to -10 ppb/K. Residues L26 and R58 despite of belonging to structural elements (β_1 strand and α-helix) shows more negative temperature coefficients indicate their involvement in weaker hydrogen bonding, hence reduced

the stability of their respective structural elements. Residues such as F20, T27, Q60 and K68 of CXCL3 and S27, Q60 of CXCL2 showed positive temperature dependence that can be attributed to the fact that these residues are experiencing the ring current effects from their neighboring aromatic residues **(Fig. 4.26 D)**. Overall, the more negative temperature coefficients observed in CXCL3 indicates that CXCL3 has a weaker hydrogen bonding pattern than CXCL2 in some structural segments, thus evidencing for its lower structural stability, which is in line with the HX studies presented in the previous section.

The alternative conformations of CXCL2 and CXCL3 were analyzed by calculating the residual curvatures of the chemical shifts. The curved temperature dependence of a residue indicates that the particular residue access more than one conformational state [38]. Further, the shape of the curvature (convex or concave) of the amide proton depends on the relative position of the residue in the structural element, and its magnitude between the native and the excited state. The origin of curved temperature dependence is the resultant of differential structural perturbations of alternate states [38]. Many of the residues in both CXCL2 and CXCL3 showed curved temperature dependence. For illustration purpose, some of the residues that exhibited convex/concave curvatures are shown in **Fig. 4.27**. All the residues showing curved temperature dependence in different structural elements have been summarized in **Table 4.8** and are marked on the sequences and structures of CXCL3 and CXCL2 **(Fig. 4.28)**. More number of curved residues were observed for CXCL3 (35) compared to CXCL2 (26) indicating the enhanced dynamic nature of the CXCL3. Around 21 residues in CXCL3 that belongs to structural elements showed curved dependence in contrast to 11 residues in CXCL2. All the common 11 residues (F20, K21, Q24, S25, V28, V40, A42, Q47, C51, I63, Q64) in CXCL2 and CXCL3 showed similar curvature profile (convex) indicating that these residues in structural elements have undergone similar alternative conformational states **(Fig.4.28 A)**. Significant differences in nature of curvatures have been observed for the residues present in the N-loop and the interconnecting loops. Out of 15 such curved residues in CXCL3, only 5 showed a similar behavior with those of CXCL2. For rest of 10 residues, CXCL2 residues exhibited an opposite type of curvature or a linear behavior. Some the residues with such same and altered curvature characteristics are depicted in **Fig. 4.27**. The higher number of residues that access low energy excited states in CXCL3, and their altered nature of curvature from its counterpart CXCL2 also establish the differential dynamic characteristics of CXCL3 **(Fig. 4.28)**.

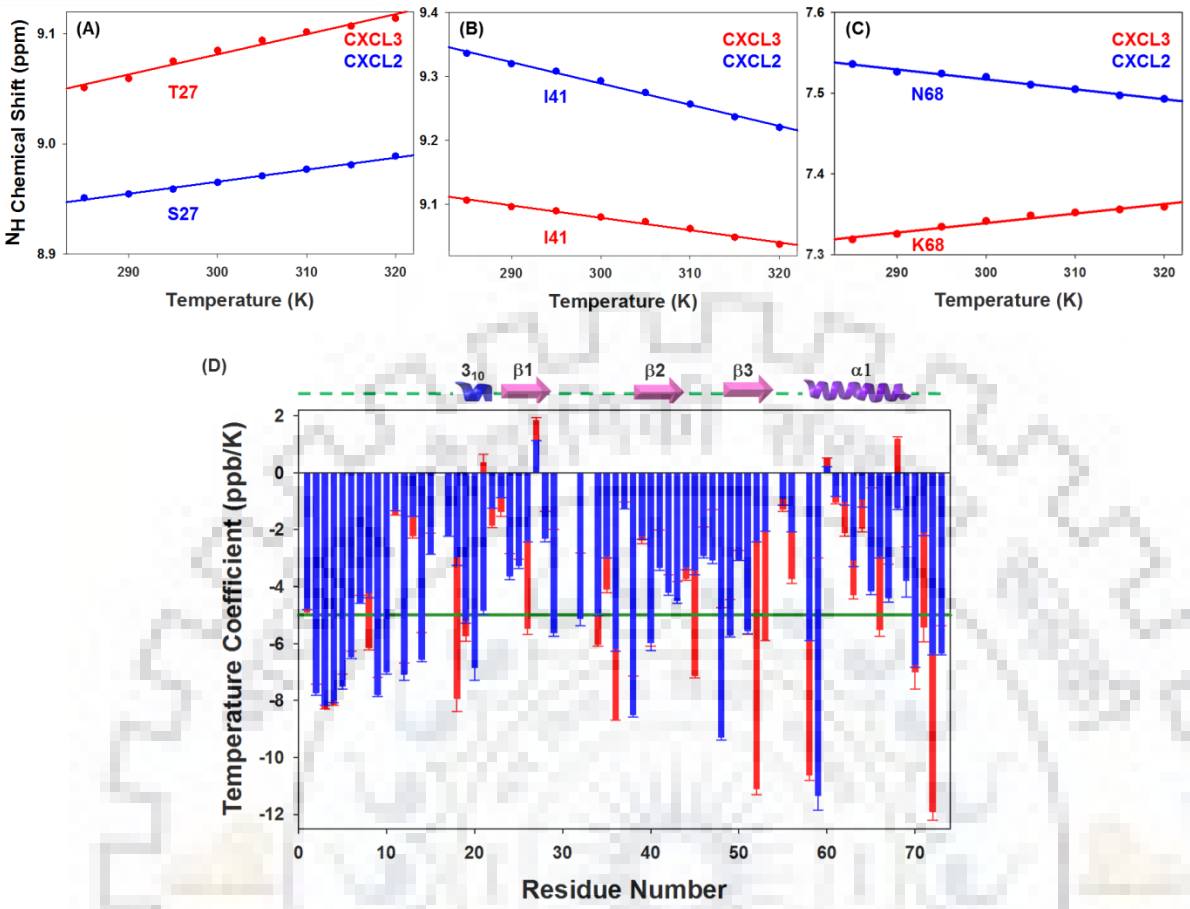


Figure 4.26: *Comparison of different profiles of temperature dependence of amide proton chemical shifts of residues in CXCL3 and CXCL2* **(A)** *S27,* **(B)** *I41, and* **(C)** *N68;* **(D)** *Overlay of temperature coefficients of CXCL3 (in red) and CXCL2 (in blue) residues. The secondary structural elements in the protein are marked above with arrows (for sheets) and helix (for helices).*

Table 4.8: *Summary of all the residues showing curved (non-linear) temperature dependence in CXCL2 and CXCL3. Numbers after the slash (/) mark indicates the total number of residues in that category.*

Residues	**CXCL2**	**CXCL3**
Curved Residues	26/73	35/73
Curved Residues in Loops	15/40	14/40
Curved Residues in structural elements	11/33	21/33
3_{10}-helix	2/3	2/3
β_1	4/6	5/6
β_2	2/6	3/6
β_3	1/6	3/6
C-terminal α-helix	2/12	8/12

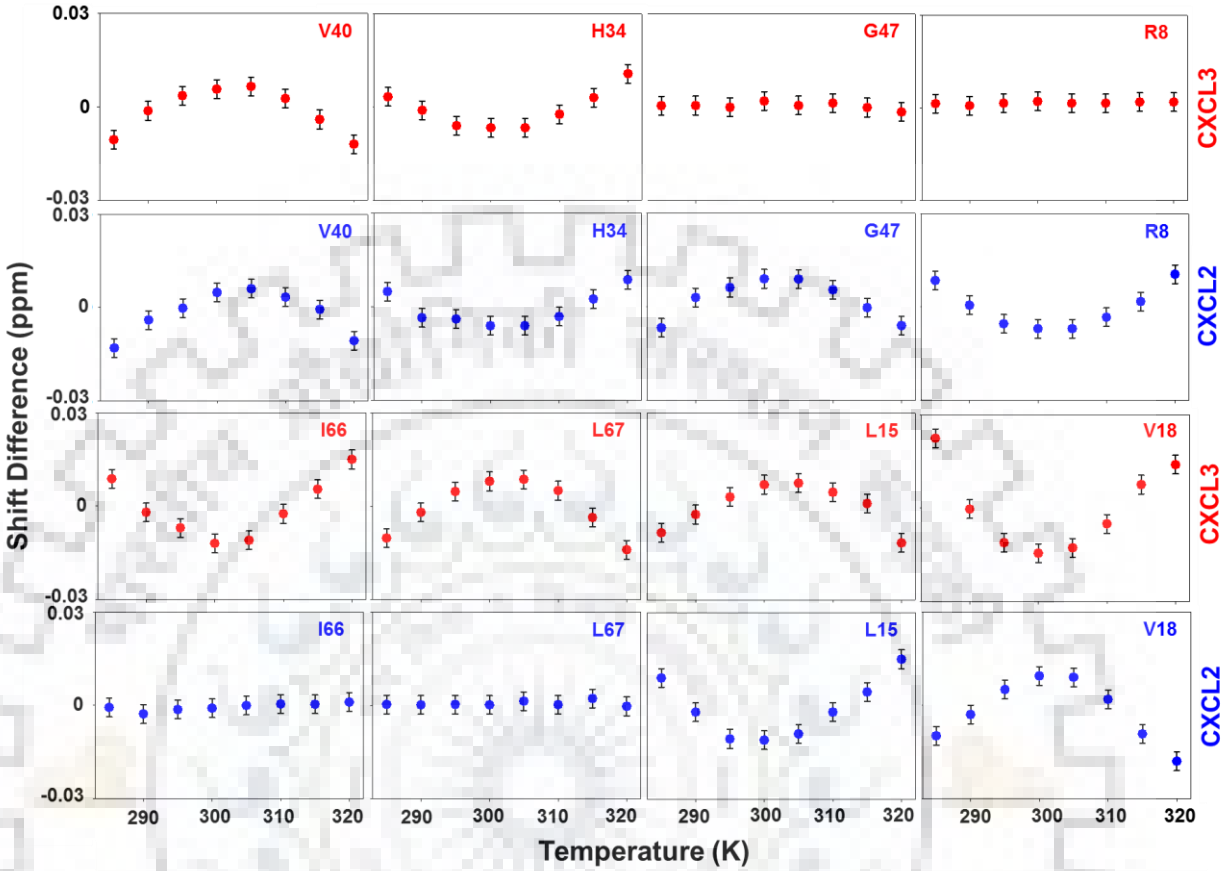


Figure 4.27: *Comparison of non-linear/linear temperature dependence profiles of amide protons of different residues of CXCL3 (in red) and CXCL2 (in blue). The chemical shifts were measured and were fitted to linear equations. The residuals obtained by subtracting the calculated value from the observed value were plotted against temperature; X-axis of each graph represents temperature range from 285 to 320K and Y-axis represents amide proton chemical shift difference (residuals) with scale of 0.06 ppm which centered at zero between +0.03 to -0.03*

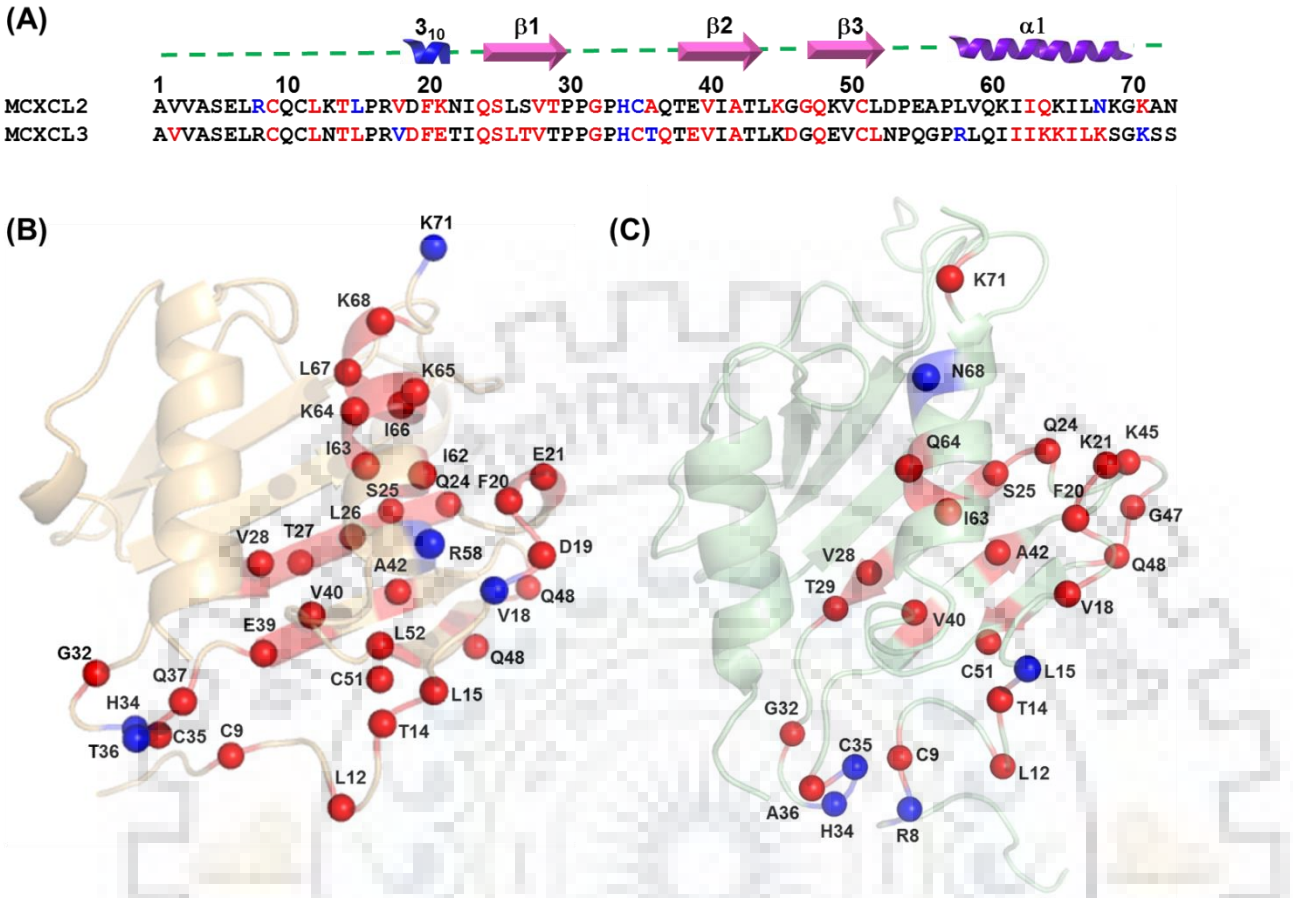


Figure 4.28: *Residues showing different temperature dependence profiles have been marked on* **(A)** *amino acid sequences of CXCL2 and CXCL3 with different colors; Linear (black), Convex (red) and Concave (blue), and as spheres of different colors (Convex- red, Concave- blue) on structure of* **(B)** *CXCL3 (modeled structure) and* **(C)** *CXCL2 (PDB: 3N52).*

4.4 Discussion

In the present chapter, murine CXCL3 chemokine was cloned and characterized, which is a pre-requisite to unravel comparative structural analysis of GRO chemokines. Structural and biophysical studies indicated the closeness of CXCL3 and CXCL2 in contrast to CXCL1. CXCL2 and CXCL3 originated as result of duplication of CXCL1, evolved and diverged from CXCL1 in terms of various properties including oligomerization, and GAG binding propensities, structural contacts. Electrostatic surface analysis of CXCL2 and CXCL3 depicted that, along with conserved surface features, these paralogs attained differentially charged surface segments to perform their specific functions. For example, they share a conserved positive surface on the α-helical side for GAG binding. However, they exhibit opposite surfaces for β-sheet as a resultant of presence of

negatively charged residues E21, D46, E49 in CXCL3 and positively charged (K21, K49) or uncharged residues (G46) in CXCL2. In addition, the presence of altering hydrophobic residues including R58, L59, I61 in CXCL3 as compared to the L58, V59, and K61 in CXCL2 marks for their differential hydrophobic surfaces. ANS binding experiments also established their differential tertiary structural features as ANS distinctly bound to CXCL2 and CXCL3. Indeed, CXCL3 possess a surface exposed hydrophobic pocket for ANS binding that is deficient in CXCL2. NMR and fluorescence data suggest that although ANS binds in the specific pocket of CXCL3 with a stoichiometry of 1:1. Specific binding of ANS (1:1) was also observed with human IL-1 RA (Interleukin-1 receptor agonist) using NMR chemical shift perturbation data, and the reported dissociation constants varied between ~ 100 μM - 300 μM [39]. Moreover, the K_d values in the range of 1-10 μM has been reported for the highly fluorescence ANS molecules from the fluorescence titration experiments on different proteins (Bacteriorhodopsin, bovine serum albumin, ovalbumin, and porcine somatotrophin) [40]. Hydrophobic pocket of ANS in CXCL3 is formed by the residues (R58, I61, I62, and F20) that are in the close vicinity to the GAG binding domain. Hence, the present study suggests that CXCL3 specific naphthalene based small molecule inhibitors can be designed. Presence of such binding pockets on specific chemokines in the vicinity of GAG binding domain provides an opportunity in designing chemokine specific small molecule inhibitors to regulate chemokine-GAG interactions/to overcome several chemokine-specific diseases or inflammatory diseases. Rational designing of such non-protein drug like small molecules is one of the challenging tasks in the current medicinal chemistry research. Naphthalene based extended inhibitor has been designed that can act as anticancer agent as it blocks the enzymatic activity of farnesoyl transferase, thereby preventing the farnesoylation of RAS oncogene [41]. Patents have also been filed for some of the novel naphthalene based inhibitors of anti-apoptotic proteins [42].

Further, NMR relaxation parameter (R_2) dictated differential dynamic features at the N-loop, 3_{10} helix, β_1 strand, and C-terminal helix of CXCL3 and CXCL2. The enhanced R_2 values observed for several structural elements (N-loop, 3_{10} helix, β_1 strand, and C-terminal helix) in CXCL3 in contrast to CXCL2 depicts their ms-μs flexibility in CXCL3. The enhanced dynamics of CXCL3 in comparison to CXCL2 is also supported by the temperature dependence experiments, where CXCL3 exhibited ~ 50 % more residues that access alternative conformations. The observed hetNOE values for CXCL3 are almost similar in all the structural elements, although few

differences were observed in N- and C- terminus. However, in contrast to the current observations, literature reports on het NOE values of disulfide bridged CXCL1 dimer suggested that the C-terminal helix possess significant flexibility in the ns-ps time scale [13]. The observed dynamics of CXCL1 C-terminal helix in comparison to CXCL3/CXCL2 can be attributed to the weak dimerization phenomenon of the former. Further, the observed relaxation parameters are in line with the dynamics parameters reported for other NAC chemokines CXCL5 and CXCL8 [43,44].

Furthermore, the enhanced dynamics can be correlated to the weaker stability of CXCL3 to that of CXCL2 as revealed by native state HX studies. The lower protection factors and the stabilization free energies of CXCL3 suggest that its quaternary packing is comparatively weaker than CXCL2. Such a variation in stabilization free energies are justifiable considering the differential surface charge patterns and the interchange of some of the hydrophobic/charged residues (K21, G46, K49, L58, V59 and K61) in CXCL2 vs (E21, D46, E49, R58, L59, I61) in CXCL3. All these features clearly establish that CXCL3 is energetically and dynamically different from CXCL2. Future quantitative receptor and GAG binding studies are essential to delineate their differential interactions with binding partners.

4.5 Conclusions

This chapter deciphered various biophysical characteristics of CXCL3 in comparison to other GRO chemokines. Biophysical studies evidenced that unlike CXCL1; CXCL3 forms tight dimer and binds to heparin with moderate affinity. The structural and stability analysis suggested that CXCL3 is thermally stable and exhibit similar structural dimeric fold as CXCL2. However, the electrostatic surface analysis and the fluorescence measurements evidenced that these closest paralogs (CXCL2 and CXCL3) possess differential charge distribution and hydrophobic pockets on their quaternary surfaces. Further, ANS binding studies confirmed the presence of surface exposed hydrophobic pocket on CXCL3. NMR relaxation and hydrogen exchange studies revealed the lower stability and highly dynamic nature of CXCL3 in comparison to CXCL2. Such distinct structural stability and dynamic features of CXCL3 and CXCL2 could possibly be the origin of their differential functional behaviors.

4.6 References

1. Haskill S, Peace A, Morris J, Sporn SA, Anisowicz A, Lee SW, Smith T, Martin G, Ralph P, Sager R. Identification of three related human GRO genes encoding cytokine functions. Proc. Natl. Acad. Sci. U. S. A. 87(19), 7732-7736 (1990).

2. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ. Defining the origins and evolution of the chemokine/chemokine receptor system. J Immunol. 176(1), 401-415 (2006).

3. Nomiyama H, Osada N, Yoshie O. The evolution of mammalian chemokine genes. Cytokine Growth Factor Rev. 21(4), 253-262 (2010).

4. Shibata K, Nomiyama H, Yoshie O, Tanase S. Genome diversification mechanism of rodent and Lagomorpha chemokine genes. Biomed. Res. Int. 2013, 856265 (2013).

5. Widdison S, Coffey TJ. Cattle and chemokines: evidence for species-specific evolution of the bovine chemokine system. Anim Genet. 42(4), 341-353 (2011).

6. Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome Biol. 7(12), 243- (2006).

7. Luan J, Shattuck-Brandt R, Haghnegahdar H, Owen JD, Strieter R, Burdick M, Nirodi C, Beauchamp D, Johnson KN, Richmond A. Mechanism and biological significance of constitutive expression of MGSA/GRO chemokines in malignant melanoma tumor progression. J Leukoc. Biol. 62(5), 588-597 (1997).

8. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, Rousseau S, Eidelman DH, Hamid Q. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. J Immunol. 191(5), 2731-2741 (2013).

9. Luttichau HR. The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist. J Biol. Chem. 285(12), 9137-9146 (2010).

10. Zhang L, Zhang L, Li H, Ge C, Zhao F *et al.* CXCL3 contributes to CD133(+) CSCs maintenance and forms a positive feedback regulation loop with CD133 in HCC via Erk1/2 phosphorylation. Sci. Rep. 6, 27426 (2016).

11. Doll D, Keller L, Maak M, Boulesteix AL, Siewert JR, Holzmann B, Janssen KP. Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. Int. J Colorectal Dis. 25(5), 573-581 (2010).

12. Otto G, Burdick M, Strieter R, Godaly G. Chemokine response to febrile urinary tract infection. Kidney Int. 68(1), 62-70 (2005).

13. Poluri KM, Joseph PR, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. J. Biol. Chem. 288(35), 25143-25153 (2013).

14. Rajasekaran D, Keeler C, Syed MA, Jones MC, Harrison JK, Wu D, Bhandari V, Hodsdon ME, Lolis EJ. A model of GAG/MIP-2/CXCR2 interfaces and its functional effects. Biochemistry. 51(28), 5642-5654 (2012).

15. Ravindran A, Sawant KV, Sarmiento J, Navarro J, Rajarathnam K. Chemokine CXCL1 dimer is a potent agonist for the CXCR2 receptor. J Biol. Chem. 288(17), 12244-12252 (2013).

16. Sawant KV, Poluri KM, Dutta AK, Sepuru KM, Troshkina A, Garofalo RP, Rajarathnam K. Chemokine CXCL1 mediated neutrophil recruitment: Role of glycosaminoglycan interactions. Sci. Rep. 6, 33123 (2016).

17. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry. 33(19), 5984-6003 (1994).

18. Peng JW, Wagner G. Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. Biochemistry. 31(36), 8571-8586 (1992).

19. Hvidt A, Nielsen SO. Hydrogen exchange in proteins. Adv. Protein Chem. 21, 287-386 (1966).

20. Bai Y, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. Science. 269(5221), 192-197 (1995).

21. Tunnicliffe RB, Waby JL, Williams RJ, Williamson MP. An experimental investigation of conformational fluctuations in proteins G and L. Structure. 13(11), 1677-1684 (2005).

22. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.4.1.

23. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput. Chem. 31(2), 455-461 (2010).

24. Fairbrother WJ, Reilly D, Colby TJ, Hesselgesser J, Horuk R. The solution structure of melanoma growth stimulating activity. J Mol. Biol. 242(3), 252-270 (1994).

25. Qian YQ, Johanson KO, McDevitt P. Nuclear magnetic resonance solution structure of truncated human GRObeta [5-73] and its structural comparison with CXC chemokine family members GROalpha and IL-8. J Mol. Biol. 294(5), 1065-1072 (1999).

26. Shao W, Jerva LF, West J, Lolis E, Schweitzer BI. Solution structure of murine macrophage inflammatory protein-2. Biochemistry. 37(23), 8303-8313 (1998).

27. Salanga CL, Handel TM. Chemokine oligomerization and interactions with receptors and glycosaminoglycans: the role of structural dynamics in function. Exp. Cell Res. 317(5), 590-601 (2011).

28. Baggiolini M. Chemokines and leukocyte traffic. Nature. 392(6676), 565-568 (1998).

29. Poluri KM. Chemokines: The Holy Messengers of Leukocyte Trafficking. Austin J Biotechnol Bioeng1(3), 3(2014).

30. Krishna Mohan PM. Unfolding energetics and conformational stability of DLC8 monomer. Biochimie. 89(11), 1409-1415 (2007).

31. Sharma M, Kumar D, Poluri KM. Elucidating the pH-Dependent Structural Transition of T7 Bacteriophage Endolysin. Biochemistry. 55(33), 4614-4625 (2016).

32. Mukherjee S, Mohan PM, Chary KV. Magnesium promotes structural integrity and conformational switching action of a calcium sensor protein. Biochemistry. 46(12), 3835-3845 (2007).

33. Collini M, D'Alfonso L, Baldini G. New insight on beta-lactoglobulin binding sites by 1-anilinonaphthalene-8-sulfonate fluorescence decay. Protein Sci. 9(10), 1968-1974 (2000).

34. Uversky VN, Winter S, Lober G. Use of fluorescence decay times of 8-ANS-protein complexes to study the conformational transitions in proteins which unfold through the molten globule state. Biophys. Chem. 60(3), 79-88 (1996).

35. 1-anilinonaphthalene-8-sulfonate (ANS); a versatile fluorescent probe from protein folding study to drug design. Biowave12, ISSN: 1598-8767 (2010).

36. Hvidt A, Nielsen SO. Hydrogen exchange in proteins. Adv. Protein Chem. 21, 287-386 (1966).

37. Baxter NJ, Williamson MP. Temperature dependence of 1H chemical shifts in proteins. J Biomol. NMR. 9(4), 359-369 (1997).

38. Williamson MP. Many residues in cytochrome c populate alternative states under equilibrium conditions. Proteins. 53(3), 731-739 (2003).

39. Latypov RF, Liu D, Gunasekaran K, Harvey TS, Razinkov VI, Raibekas AA. Structural and thermodynamic effects of ANS binding to human interleukin-1 receptor antagonist. Protein Sci. 17(4), 652-663 (2008).

40. Cardamone M, Puri NK. Spectrofluorimetric assessment of the surface hydrophobicity of proteins. Biochem. J. 282 (Pt 2), 589-593 (1992).

41. Burns CJ, Guitton JD, Baudoin B, Lelievre Y, Duchesne M, Parker F, Fromage N, Commercon A. Novel conformationally extended naphthalene-based inhibitors of farnesyltransferase. J Med. Chem. 40(12), 1763-1767 (1997).

42. Maurizio Pellecchia and John C.Reed. Naphthalene-based inhibitors of anti-apoptotic proteins. Burnham Institute For Medical ResearchUS 9115061 B2 (2015).

43. Sepuru KM, Poluri KM, Rajarathnam K. Solution structure of CXCL5--a novel chemokine and adipokine implicated in inflammation and obesity. PLoS. One. 9(4), e93228- (2014).

44. Grasberger BL, Gronenborn AM, Clore GM. Analysis of the backbone dynamics of interleukin-8 by 15N relaxation measurements. J Mol. Biol. 230(2), 364-372 (1993).

Chapter 5: Concluding Remarks and Future Perspectives

5.1 Concluding remarks

In the present work, the evolutionary history of GRO genes across the diverse range of mammals was reconstructed. A trend of species specific evolution pattern has been followed by GRO chemokines. Although these genes underwent purifying selection, 17 positively selected sites majorly residing in the GAG binding segments were observed. Some of these sites have coevolved as a consequence of their role in structural stabilities. GRO genes encountered selection pressures in species dependent manner to develop their specific functional superiorities. Moreover, alterations in their nucleotide/amino acid sequences in particular, resulted in the emergence of species specific electrostatic surfaces. As a resultant, GRO genes exhibit different geometries/modes of GAG binding. These observations demonstrate that the evolutionary programming of GRO genes is coupled with their surface modulations and functional variance.

Considering the fact that GRO genes evolved differentially, a comparative study of murine GRO chemokines was carried out. Homo/Hetero oligomerization and GAG binding properties of GRO chemokines (CXCL1 and CXCL2) for which the structures are available established that the CXCL1 and CXCL2 form homo and hetero oligomers with differential efficacies under in-vitro conditions. Such phenomenon of chemokine oligomerization is very important for their variable functionality and therapeutic formulation, as monomeric and dimeric forms will generate differential gradients. Further, GAG binding studies evidenced that GAG binding induces both homo and heterodimerization of CXCL1 and CXCL2 chemokines. The sulfation pattern of GAGs plays an important role in defining their interaction with chemokines and henceforth their oligomerization. Both HP6 and SHA6 induced homo-/hetero-dimerization in CXCL1 and CXCL2 in contrast to HA6 and NC6. The strong binding nature of sulfated hyaluronan (SHA6) suggests that this synthetic molecule can potentially compete with natural GAG heparin/heparan sulfate. Hence, SHA6 can be exploited in formulating GAG based therapeutics for regulating neutrophil trafficking.

After delineating the oligomerization and GAG binding properties of GRO chemokines CXCL1 and CXCL2, the oligomerization, and structural properties of third member of GRO chemokine family (CXCL3) were studied. The oligomerization and heparin binding characteristics were compared with its paralogs (CXCL1/CXCL2). These studies evidenced oligomeric and GAG (heparin) binding properties of CXCL3 are similar to CXCL2 and are far different from CXCL1. However, CXCL3 possess differential surface charges and hydrophobic pockets compared to CXCL2 as evident from fluorescence and electrostatic surface analysis. ANS binding studies established that CXCL3 contains a specific surface exposed hydrophobic pocket, and binds ANS with a stoichiometry of 1:1 per CXCL3 monomer. NMR studies including relaxation parameters, native state hydrogen exchange stabilities, and amide proton temperature dependence, pointed towards the higher flexibility and low stability features of CXCL3 in contrast to its paralog CXCL2. Such distinct stability and oligomerization features of these duplicated genes in conjunction to their conserved dimeric fold deciphered the underlying mechanisms for their common and protein specific functions, thus throwing light on their evolution-structure-stability-function relationships.

5.2 Future perspectives

Taking into account the observed evolutionary patterns of GRO chemokines, it is imperative to further explore evolutionary perspectives for more number of laurasiatherian species in order to throw light on ancestral origin of GRO gene duplication. Such information is possible upon complete gene annotation for various species under this branch. These evolution-structure relationship studies have to be expanded to neutrophil activating chemokines to decode their evolutionary relation with the GRO chemokines across various species. Further, to get insights into the basis of their differential biological functions, more comprehensive structural and molecular level studies of GRO chemokines with their cellular binding partners (receptor and GAGs) are essential. Molecular studies for GRO chemokines from other species including human and horse is essential to unwind their species specific properties and behaviors with cellular partners as it has been observed that human and horse GRO chemokines exhibit multiple GAG binding surfaces. Further, studies with specific sulfated GAGs such as chondroitin 6-sulfate and chondroitin 4-sulfate will yield more information about the specificity of sulfate groups for chemokine binding. It is also vital to draw the molecular level details about receptor

binding of GRO chemokines, as one of the functional study reported that CXCL1 activates only one receptor (CXCR2) whereas CXCL2 and CXCL3 activates both the receptors (CXCR1 and CXCR2).

Understanding the mechanistic details of their biological functions will aid in designing engineered chemokines as novel protein based therapeutics. Engineered monomers and dimers of GRO chemokines will help in dissecting their differential stability, dynamic features, and role of dimerization in regulating their functional characteristics. To obtain comprehensive insights into the molecular mechanisms, atomic level high resolution structures of GRO complexes with receptors and GAGs with both monomeric and dimeric proteins are imperative. Molecular insights into this complex network of interactions will aid us in formulating chemokine variants with altered biological properties as therapeutics for a number of inflammatory and infectious diseases. As the present study established the presence of ANS binding pocket in CXCL3 in contrast to CXCL2, hence the CXCL3 specific naphthalene based small molecule inhibitors can be designed. Identifying such binding pockets on specific chemokines provides an opportunity to design chemokine specific inhibitors to regulate chemokine-GAG/receptor interactions in order to resolve several chemokine mediated inflammatory diseases.

Appendices

Appendix-I

Nucleotide sequence IDs of GRO family chemokines (CXCL1, CXCL2, and CXCL3) from different mammalian species.

Species	CXCL1	CXCL2	CXCL3
African bush elephant	-	-	XM_010594141.1
Bactrian camel	XM_010969408.1		
Baiji	XM_007463218.1	XM_007463219.1	-
Bison	XM_010838198.1	-	XM_010838177.1
Black-capped squirrel monkey	XM_010342726.1	XM_010342728.1	-
Bovine	NM_175700.1	NM_001048165.1	NM_001046513.2
Brown rat	NM_030845.1	NM_053647.1	NM_138522.1
Cape golden mole	XM_006871594.1	-	-
Cat	XM_011279689.1	-	-
Cape elephant shrew	XM_006898468.1	-	-
Chinese hamster	NM_001244044.1	XM_007630617.1	XM_001244139.1
Chimpanzee	XM_001156094.4	XM_001155614.4	XM_517228.4
Chinese tree shrew	XM_006142920.2	-	-
Crab-eating macaque	AB262775.1	AB262776.2	AB262777.1
Common bottlenose dolphin	XM_004319600.1	-	-
Deer mouse	-	XM_006993515.2	-
Ferret	XM_004766349.2		
Gibbon	XM_012499279.1	ENSNLET00000010136	XM_003265742.2
Giant panda	XM_002919144.2	-	-
Gorilla	XM_004038813.1	XM_004038819.1	-
Goat	XM_013964688.1		
Golden hamster	-	XM_005068086.2	-
Guinea pig	NM_001172938.1	-	-
Gray short-tailed opossum	XM_007495669.2	-	-
Hedgehog	-	-	XM_004703375.1
House mouse	NM_008176.3	NM_009140.2	NM_203320.3
Human	NM_001511.3	NM_002089.3	NM_002090.2
Horse	NM_001309480.1	NM_001143955.1	NM_001143793.2
Killer whale	XM_012538162.1	-	-
Little brown bat	-	-	XM_006094377.2
Marmoset	XM_002745753.3	-	-
Mouflon	XM_012135521.2	-	-
Minke whale	XM_007179836.1	-	-
Nine-banded armadillo	-	-	XM_004465324.2
Naked mole rat	XM_004833919.1	-	-
Orangutan	XM_002814861.3	XM_002814867.3	XM_002814865.2
Ord's kangaroo rat	XM_013020161.1	-	-
Philipine tarsier	XM_008059500.1	-	-
Pig	NM_001001861.2	NM_001001861.2	XM_005666754.2
Platypus	XM_007669225.1	-	-
Rabbit	U95808.1	ENSOCUT00000031529	U12310.1
Rhesus macaque	NM_001032878.1	- -	NM_001032879.1
Sperm whale	XM_007126262.1	-	-
Star nosed mole	-	-	XM_004681217.2
Thirteen-lined GROund squirrel	-	XM_005333294.2	XM_005333245.1
Tasmanian devils	XM_012551956.1	-	-
Water buffalo	XM_006059633.1	XM_006059634.1	-
Walrus	-	XM_004392914.1	XM_004392941.1
West Indian manatee	-	-	XM_004383312.1
White rhinoceros	-	XM_004419171.2	XM_004419172.2
Yak	-	-	XM_005891250.2

Appendix-II

Ramachandran plot statistics for modeled murine / human / horse CXCL (GRO) chemokine structures.

	Murine CXCL1		Murine CXCL3		Human CXCL3	
Residues in most favoured regions [A,B,L]	108	88.5%	107	93.9%	93	78.8%
Residues in additional allowed regions [a,b,l,p]	13	10.7%	6	5.3%	25	21.2%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.8%	1	0.9%	0	0.0%
Residues in disallowed regions	0	0.0%	0	0.0%	0	0.0%
Number of non-glycine and non-proline residues	122	100.0%	114	100.0%	118	100.0%
Number of end-residues (excl. Gly and Pro)	2		4		4	
Number of glycine residues (shown as triangles)	10		8		8	
Number of proline residues	12		12		8	
Total number of residues	146		138		138	

	Horse CXCL1		Horse CXCL2		Horse CXCL3	
Residues in most favoured regions [A,B,L]	110	92.45%	110	92.4%	110	92.4%
Residues in additional allowed regions [a,b,l,p]	8	6.7%	8	6.7%	8	67.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.8%	1	0.8%	1	0.8%
Residues in disallowed regions	0	0.0%	0	0.0%	0	0.0%
Number of non-glycine and non-proline residues	119	100.0%	119	100.0%	119	100.0%
Number of end-residues (excl. Gly and Pro)	3		3		3	
Number of glycine residues (shown as triangles)	8		8		8	
Number of proline residues	7		7		7	
Total number of residues	137		137		138	